



US007846454B2

(12) **United States Patent**
Yanagi et al.

(10) **Patent No.:** **US 7,846,454 B2**
(45) **Date of Patent:** ***Dec. 7, 2010**

(54) **CLONED GENOMES OF INFECTIOUS
HEPATITIS C VIRUSES AND USES THEREOF**

WO WO 98/39031 9/1998

(75) Inventors: **Masayuki Yanagi**, Rockville, MD (US);
Jens Bukh, Bethesda, MD (US);
Suzanne U. Emerson, Rockville, MD
(US); **Robert H. Purcell**, Boyds, MD
(US)

(73) Assignee: **The United States of America as
represented by the Department of
Health and Human Services**,
Washington, DC (US)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 246 days.

This patent is subject to a terminal dis-
claimer.

(21) Appl. No.: **11/784,809**

(22) Filed: **Apr. 9, 2007**

(65) **Prior Publication Data**

US 2008/0014212 A1 Jan. 17, 2008

Related U.S. Application Data

(62) Division of application No. 09/662,454, filed on Sep.
14, 2000, now Pat. No. 7,201,911, which is a division
of application No. 09/014,416, filed on Jan. 27, 1998,
now Pat. No. 6,153,421.

(60) Provisional application No. 60/053,062, filed on Jul.
18, 1997.

(51) **Int. Cl.**

A61K 39/12 (2006.01)

A61K 39/29 (2006.01)

C12N 15/00 (2006.01)

C12N 15/09 (2006.01)

C12N 7/00 (2006.01)

C12N 15/51 (2006.01)

C12N 5/10 (2006.01)

(52) **U.S. Cl.** **424/199.1**; 424/202.1; 424/205.1;
435/455; 435/456; 435/91.41; 435/91.42

(58) **Field of Classification Search** None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,679,342 A 10/1997 Houghton et al.
6,127,116 A * 10/2000 Rice et al. 435/6
6,153,421 A 11/2000 Yanagi et al.
7,084,266 B1 8/2006 Yanagi et al.
7,201,911 B1 * 4/2007 Yanagi et al. 424/218.1
2007/0141668 A1 6/2007 Yanagi et al.

FOREIGN PATENT DOCUMENTS

EP 0 516 270 A2 12/1992
WO WO 97/08310 3/1997

OTHER PUBLICATIONS

Bassett, S.E. et al. (2001) "Protective immune response to hepatitis C
virus in chimpanzees rechallenged following clearance of primary
infection." *Hepatology* 33:1479-1487.

Bigger, C.B. et al. (2001) "DNA microarray analysis of chimpanzee
liver during acute resolving hepatitis C virus infection." *J. Virol.*
75(15):7059-7066.

Bukh, J. et al. (2004) "Persistence of HCV after homologous
monoclonal re-challenge is associated with emergence of virus vari-
ants." In Jilbert AR, Grgacic EVL, Vickery K, Burrell CJ, Cossart YE,
eds. *Proceedings of the 11th International Symposium on Viral Hepa-
titis & Liver Disease*. Australian Centre for Hepatitis Virology, pp.
375-377.

Dash, S. et al. (1997) "Transfection of HepG2 cells with infectious
hepatitis C virus genome." *Am. J. Pathol.* 151:363-373.

Farci, P. et al. (1996) "Prevention of hepatitis C virus infection in
chimpanzees by hyperimmune serum against the hypervariable
region 1 of the envelope 2 protein." *PNAS USA* 93:15394-15399.

Fausto, N. (1997) "Replication of hepatitis C virus in tissue culture:
a potential breakthrough in need of confirmation." *Am. J. Pathol.*
151:361.

Forns, X. et al. (2002) "The challenge of developing a vaccine against
hepatitis C virus." *J. Hepatol.* 37:684-695.

Honda, M. et al. (1996) Structural requirements for initiation of
translation by internal ribosome entry within genome-length hepati-
tis C virus RNA. *Virology* 222:31-42, Aug. 1996.

Inchauspe, G. et al. (1991) "Genomic structure of the human proto-
type strain H of hepatitis C virus: comparison with American and
Japanese isolates." *PNAS USA* 88:10292-10296.

Kolykhalov, A.A. et al. (1996) "Identification of a highly conserved
sequence element at the 3' terminus of hepatitis C virus genome
RNA." *J. Virol.* 70, 3363-3371.

Kolykhalov, A.A. et al. (1997) "Transmission of hepatitis C by
intrahepatic inoculation with transcribed RNA." *Science* 277:570-
574.

Kolykhalov, A.A. et al. (2000) "Hepatitis C virus-encoded enzymatic
activities and conserved RNA elements in the 3' nontranslated region
are essential for virus replication in vivo." *J. Virol.* 74:2046-2051.

Lanford, R.E. et al. (2004) "Cross-genotype immunity to hepatitis C
virus." *J. Virol.* 78(3):1575-1581.

Lindenbach, B.D. et al. (2006) "Cell culture-grown hepatitis C virus
is infectious in vivo and can be recultured in vitro." *PNAS USA*
103:3805-3809.

Major, M.E. et al. (2002) "Previously infected and recovered chim-
panzees exhibit rapid responses that control hepatitis C virus repli-
cation upon rechallenge." *J. Virol.* 76(13):6586-6595.

(Continued)

Primary Examiner—Maria B Marvich

(74) *Attorney, Agent, or Firm*—Townsend and Townsend and
Crew LLP

(57) **ABSTRACT**

The present invention discloses nucleic acid sequences which
encode infectious hepatitis C viruses and the use of these
sequences, and polypeptides encoded by all or part of these
sequences, in the development of vaccines and diagnostics for
HCV and in the development of screening assays for the
identification of antiviral agents for HCV.

37 Claims, 49 Drawing Sheets

OTHER PUBLICATIONS

- Major, M.E. et al. (2004) "Hepatitis C virus kinetics and host responses associated with disease and outcome of infection in chimpanzees." *Hepatology* 39(6):1709-1720.
- Meunier J.C. et al. (2005) "Evidence for cross-genotype neutralization of hepatitis C virus pseudo-particles and enhancement of infectivity by apolipoprotein C1." *PNAS USA* 102(12):4560-4565.
- Ogata, N. et al. (1991) "Nucleotide sequence and mutation rate of the H strain of hepatitis C virus." *PNAS USA* 88:3392-3396.
- Okamoto, H. et al. (1992) "Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability." *Virology* 190:894-899.
- Okamoto, H. et al. (1994) "Superinfection of chimpanzees carrying hepatitis C virus of genotype II/1b with that of genotype III/2a or I/1a." *Hepatology* 20:1131-1136.
- Rice, C.M. (2006) "Hepatitis C: progress toward new therapies from Third Annual PHRI Symposium.", published 2006 and downloaded Jan. 23, 2006.
- Su, A. et al. (2002) "Genomic analysis of the host response to hepatitis C virus infection." *PNAS USA* 99(24):15669-15674.
- Thimme, R. et al. (2002) "Viral and immunological determinants of hepatitis C virus clearance, persistence, and disease." *PNAS USA* 99(24):15661-15668.
- Weiner, A.J. et al. (2001) "Intrahepatic genetic inoculation of hepatitis C virus RNA confers cross-protective immunity." *J. Virol.* 75(15):7142-7148.
- Yanagi, M. et al. (1997) "Transcripts from a single full-length cDNA clone of Hepatitis C virus are infectious when directly transfected into the liver of a chimpanzee." *PNAS USA* 94: 8738-8743.
- Yanagi, M. et al. (1998) "Transcripts of a chimeric cDNA clone of hepatitis C virus genotype 1b are infectious in vivo." *Virology* 244: 161-172 Apr. 25.
- Yanagi, M. et al. (1999) "Hepatitis C virus: An infectious molecular clone of a second major genotype (2a) and lack of viability of intertypic 1a and 2a chimeras." *Virology* 262(1):250-263.
- Yoo, B.J. et al. (1995) "Transfection of a differentiated human hepatoma cell line (Huh 7) with in vitro-transcribed hepatitis C virus (HCV) RNA and establishment of a long-term culture persistently infected with HCV." *J. Virol.* 69:32-38, Jan. 1995.
- Yu, M.-YW. et al. (2004) "Neutralizing antibodies to hepatitis C virus (HCV) in immune globulins derived from anti-HCV-positive plasma." *PNAS USA* 101(20):7705-7710.

* cited by examiner

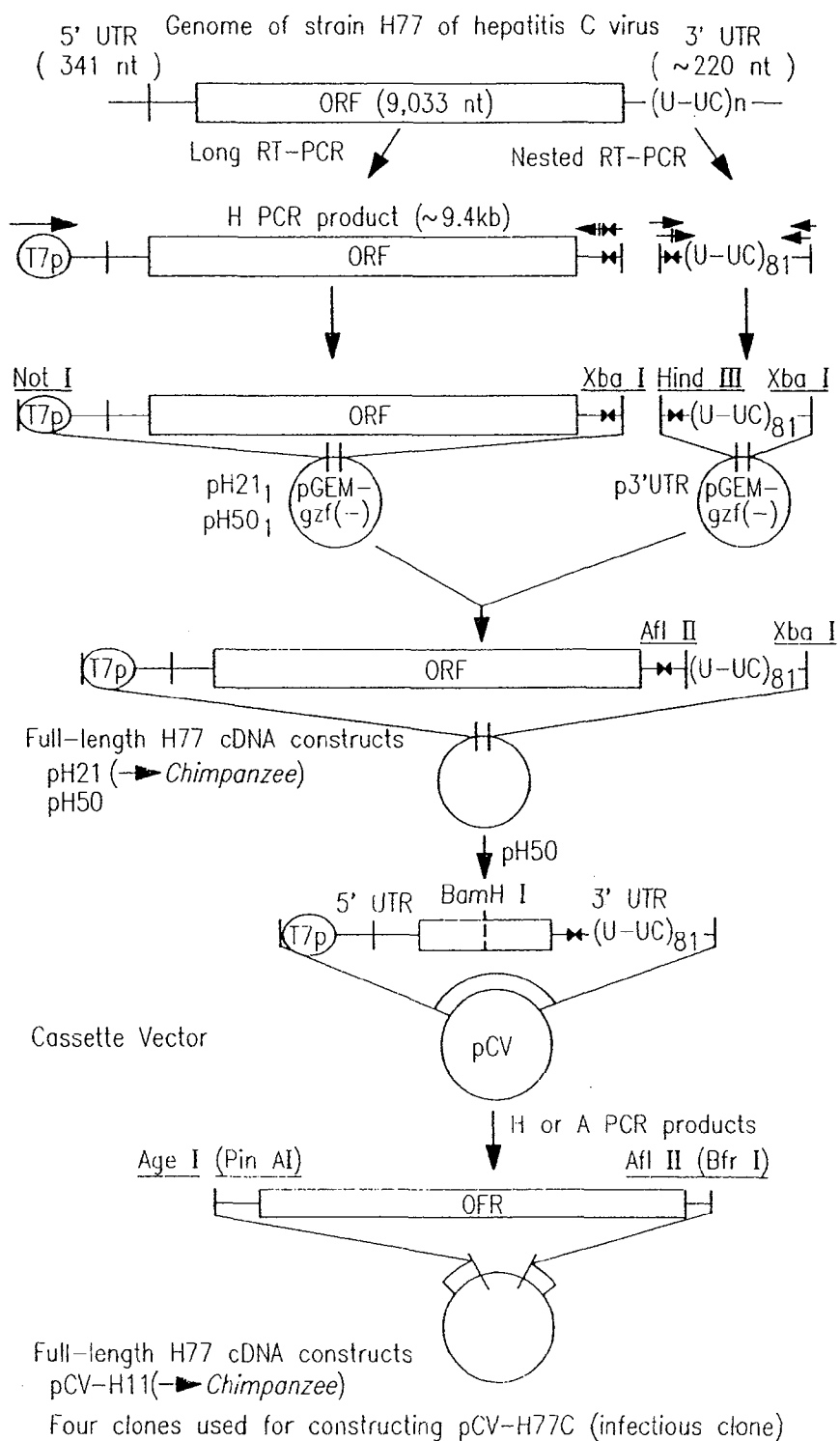
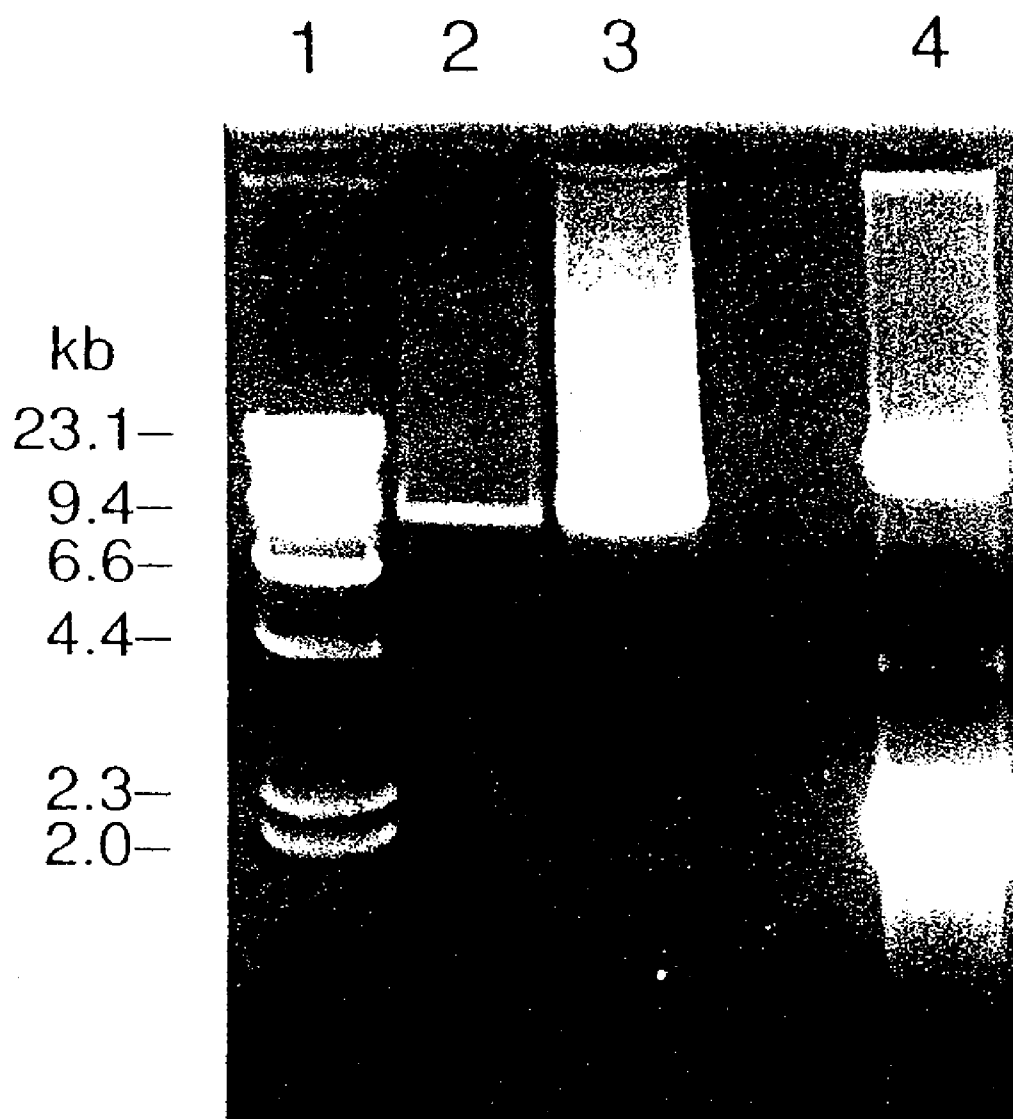


FIG. 1

**FIG. 2**

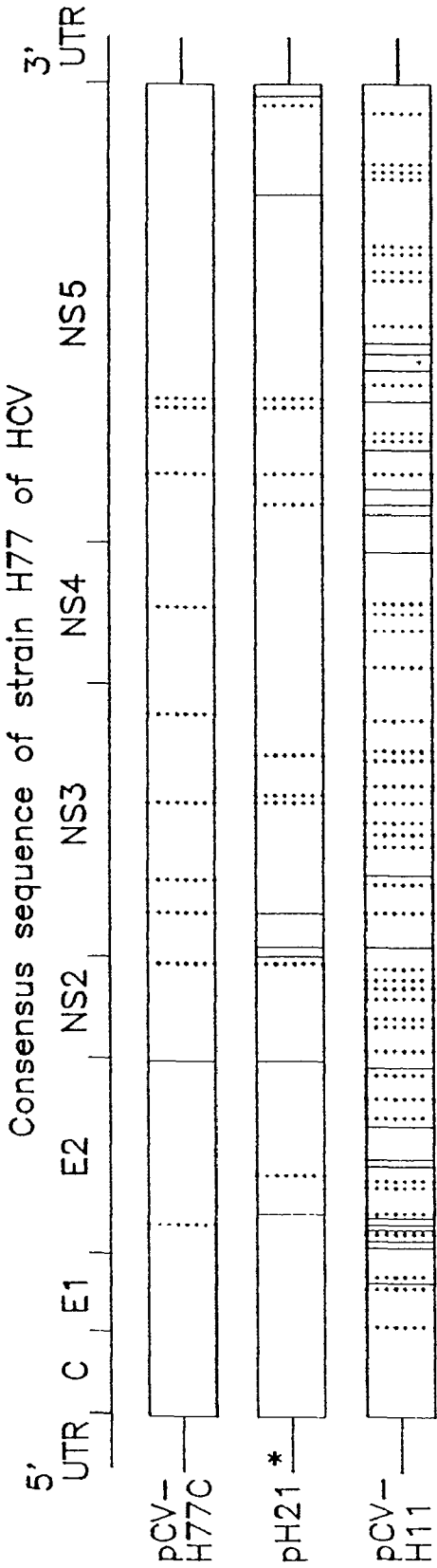


FIG. 3

H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GCCAGCCCC	TGATGGGGC	GACACTCCAC	CATGAATCAC	TCCCCGTGTA	50
GGAACTACTG	TCTTCACGA	GAAAGOGTCT	AGCCATGGCG	TTAGTATGAG	100
TGTGTTGCAG	CCTCCAGGAC	CCCCCTCCC	GGGAGAGCCA	TAGTGGTCTG	150
CGGAACCGGT	GAGTACACCG	GAATTGCCAG	GACGACCGGG	TCCTTTCTTG	200
GATAAACCCG	CTCAATGOCT	GCAGATTITG	GCGTGCCCC	GCAAGACTGC	250
TAGCCGAGTA	GTGTTGGGTC	GCGAAAGGCC	TTGTGGTACT	GCCTGATAGG	300
GTGCTTGCGA	GTGCCCCGGG	AGGTCCTGTA	GACCGTGCAC	CATGAGCACG	350
AATCCTAAAC	CTCAAAGAAA	AACCAAAAGT	AACAACCAACC	GTGCCCCACA	400
GGACGTCAAG	TTCCCCGGTG	GCGGTACAGAT	CGTTGGTGGG	GTTTACTTGT	450
TGCGCGCAG	GGGCCCCAGA	TTGGGTGTGC	GCGCGACGAG	GAAGACTTCC	500
GAGCGGTGCG	AACCTCGAGG	TAGACGTACG	CCTATCCCCA	AGGCACGTGCG	550
GCCCCAGGGC	AGGACCTGGG	CTCAGCCCCG	GTACCCCTTG	CCCCCTCTATG	600
GCAATGAGGG	TTGCGGGTGG	GCGGGATGGC	TCTGTCTCTC	CCGTGGCTCT	650
CGGCTAGCT	GGGGCCCCAC	AGACCCCCGG	CGTAGGTGCG	GCAATTTGGG	700
TAAGGTCATC	GATACCCCTA	CGTGCGGCTT	CGCCGACCTC	ATGGGGTACA	750
TACCGCTCGT	CGGCGCCCCCT	CTTGAGAGCG	CTGCCAGGGC	CCTGGCGCAT	800
GGCGTCCGGG	TTCTGGAAGA	CGGCGTGAAC	TATGCAACAG	GGAACCTTCC	850
TGGTTGCTCT	TTCTCTATCT	TCCTTCTGGC	CCTGCTCTCT	TGCTGACTG	900
TGCCCCCTTC	AGCCTACCAA	GTGCGCAATT	CCTCGGGGCT	TTACCATGTC	950
ACCAATGATT	GCCCTAACTC	CAGTATTGTG	TACGAGGCGG	CCGATGCCAT	1000
CCTGCACACT	CCGGGGTGTG	TCCCTTGGGT	TGCGGAGGGT	AACGCTCGA	1050
GGTGTGTTGGT	GGCGGTGACC	CCCACGGTGG	CCACCAGGGA	CGGCAAACTC	1100
CCCACAACGC	AGCTTGCAGC	TCATATCGAT	CTGCTTGTGCG	GGAGCGCCAC	1150
CCTCTGCTCG	GCCCTCTACG	TGGGGGACCT	GTGCGGGTCT	GTCTTTCTTG	1200
TTGGTCAACT	GTTTACCTTC	TCTOCCAGGC	GCCACTGGAC	GACGCAAGAC	1250
TGCAATTGTT	CTATCTATCC	CGGCCATATA	AOGGGTCATC	GCATGGCATG	1300
GGATATGATG	ATGAACTGGT	CCCCACGGC	AGCGTTGGTG	GTAGCTCAGC	1350
TGCTCCGGAT	CCCACAAGCC	ATCATGGACA	TGATCGCTGG	TGCTCACTGG	1400
GGAGTCCCTG	CGGCCATAGC	GTATTTCTCC	ATGGTGGGGA	ACTGGGCGAA	1450
GGTCCCTGGTA	GTCCTGCTGC	TATTTTGGCG	CGTCGACGGG	GAAACCCACG	1500
TCAACCGGGG	AAATGCCGGC	CGCACCAAGG	CTGGGCTTGT	TGGTCTCCTT	1550
ACACCAGGCG	CCAAGCAGAA	CATCCAACCTG	ATCAACACCA	ACGGCAGTTG	1600
GCACATCAAT	AGCACGGCCT	TGAATTGCAA	TGAAAGCCTT	AACACCGGCT	1650
GGTTAGCAGG	GCTCTTCTAT	CAACACAAAT	TCAACTCTTC	AGGCTGTCTT	1700
GAGAGGTTGG	CCAGCTGCCG	ACGCTTACC	GATTTTGGCC	AGGGCTGGGG	1750
TCCTATCAGT	TATGCCAACG	GAAGCGGCTT	CGACGAACGC	CCCTACTGCT	1800
GGCACTACCC	TCCAAGACCT	TGTGGCATTG	TGCCCCGAAA	GAGCGTGTGT	1850
GGCCCGGTAT	ATTGCTTCAC	TCCCAGCCCC	GTGGTGGTGG	GAACGACCGA	1900

FIG. 4A

H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CAGGTCGGGC	GCGCCTACCT	ACAGCTGGGG	TGCAAATGAT	ACGGATGTCT	1950
TCGTCTTAA	CAACACCAGG	CCACCGCTGG	GCAATTGGTT	CGGTTGTACC	2000
TGGATGAACT	CAACTGGATT	CACCAAAGTG	TGCGGAGGCG	CCCCTTGTGT	2050
CATCGGAGGG	GTGGGCAACA	ACACCTTGCT	CTGCCCCACT	GATTGCTTCC	2100
GCAAACATCC	GGAAGCCACA	TACTCTGGGT	GCGGCTCCGG	TCCCTGGATT	2150
ACACCCAGGT	GCAITGGTGA	CTACCCGAT	AGGCTTTGGC	ACTATCCTTG	2200
TACCATCAAT	TACAACATAT	TCAAAGTCAG	GATGTACGTG	GGAGGGGTGG	2250
AGCACAGGCT	GGAAGCGGCG	TGCAACTGGA	CGCGGGGCGA	AACCTGTGAT	2300
CTGGAAGACA	GGGACAGGTC	CGAGCTCAGC	CCGTTGCTGC	TGTCCACCAC	2350
ACAGTGGCAG	GTCTTCCGT	GTTCTTTTAC	GACCTTGCCA	GCCTTGTCCA	2400
CCGGCCTCAT	CCACCTCCAC	CAGAACATTG	TGGACGTGCA	GTACTTGTAC	2450
GGGGTAGGGT	CAAGCATGCG	GTCTGGGGCC	ATTAAAGTGGG	AGTACGTCTT	2500
TCTCCTGTTC	CTTCTGCTTG	CAGACGCGCG	CGTCTGCTCC	TGCTTGTGGA	2550
TGATGTACT	CATATCCCAA	GCGGAGGCGG	CTTTGGAGAA	CCTCGTAATA	2600
CTCAATGCAG	CATCCCTGGC	CGGGACGCAC	GGTCTTGTTT	CCTTCCTCGT	2650
GTTCTTCTGC	TTTGCGTGGT	ATCTGAAGGG	TAGGTGGGTG	CCCGGAGCGG	2700
TCTACGCCCT	CTACGGGATG	TGGCCTCTCC	TCTGCTCCT	GCTGGCGTTG	2750
CCTCAGCGGG	CATACGCACT	GGACACCGAG	GTGGCCCGGT	CGTGTGGCGG	2800
CGTTGTTCTT	GTGGGGTTAA	TGGCGCTGAC	TCTGTGCGCA	TATTACAAGC	2850
GCTATATCAG	CTGGTGCATG	TGGTGGCTTC	AGTATTTTCT	GACCAGAGTA	2900
GAAGCGCAAC	TGCAAGTGTG	GGTTCCCCCC	CTCAACGTCC	GGGGGGGGCG	2950
CGATGCCGTC	ATCTTACTCA	TGTGTGTAGT	ACACCCGACC	CTGGTATTTG	3000
ACATCACCAA	ACTACTCCTG	GCCATCTTCC	GACCCCTTTG	GATTCTTCAA	3050
GCCAGTTTGC	TTAAAGTCCC	CTACTTGTGG	CGCGTTCAAG	GCCTTCTCCG	3100
GATCTGCGCG	CTAGCGCGGA	AGATAGCGGG	AGGTCATTAC	GTGCAAATGG	3150
CCATCATCAA	GTTAGGGGGG	CTTACTGGCA	CCTATGTGTA	TAACCATCTC	3200
ACCCCTCTTC	GAGACTGGGC	GCACAACGGC	CTGCGAGATC	TGGCCGTGGC	3250
TGTGGAACCA	GTCGTCTTCT	CCCGAATGGA	GACCAAGCTC	ATCAAGTGGG	3300
GGGCAGATAC	CGCCCGGTGC	GGTGACATCA	TCAACGGCTT	GCCCGTCTCT	3350
GCCCGTAGGG	GCCAGGAGAT	ACTGCTTGGG	CCAGCCGACG	GAATGGTCTC	3400
CAAGGGGTGG	AGGTTGCTGG	CGCCCATCAC	GGCGTACGCC	CAGCAGACGA	3450
GAGGCCCTCT	AGGGTGTATA	ATCACCAGCC	TGACTGGCCG	GGACAAAAAC	3500
CAAGTGGAGG	GTGAGGTCCA	GATCGTGTCA	ACTGCTACCC	AAACCTTCTT	3550
GGCAACGTGC	ATCAATGGGG	TATGCTGGAC	TGTCTACCAC	GGGGCCGGAA	3600
CGAGGACCAT	CGCATCACCC	AAGGGTCTTG	TCATCCAGAT	GTATACCAAT	3650
GTGGACCAAG	ACCTTGTGGG	CTGGCCCGCT	CCTCAAGGTT	CCCGCTCAAT	3700
GACACCTGT	ACCTGCGGCT	CCTCGGACCT	TTACCTGGTC	ACGAGGCACG	3750
CCGATGTCAT	TCCCGTGGCG	CGCGGAGGTG	ATAGCAGGGG	TAGCCTGCTT	3800

FIG. 4B

H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TGCCCCCGGC	CCATTTCCTA	CTTGAAAGGC	TCTTCGGGGG	GTCCGCTGTT	3850
GTGCCCCGCG	GGACACGCG	TGGGCTATT	CAGGGCCGCG	GTTGCAACC	3900
GTTGAGTGGC	TAAAGCGGTG	GACTTTATCC	CTGTGGAGAA	CCTAGGGACA	3950
ACCATGAGAT	CCCCGGTGTT	CAOGGACAAC	TCTCTCCAC	CAGCAGTGCC	4000
CCAGAGCTTC	CAGGTGGCCC	ACCTGCATGC	TCCCACCGGC	AGCGGTAAAG	4050
GCACCAAGGT	CCCGGCTGG	TACGCAGCCC	AGGGCTACAA	GGTGTGGTG	4100
CTCAACCCCT	CTGTTGCTGC	AAGGCTGGGC	TTTGGTGCTT	ACATGTCCAA	4150
GGCCCATGGG	GTTGATCCTA	ATATCAGGAC	CGGGGTGAGA	ACAATTACCA	4200
CTGGCAGCCC	CATCACGTAC	TCCACTACG	GCAAGTTCTT	TGCGGACCGC	4250
GGGTGCTCAG	GAGGTGCTTA	TGACATAATA	ATTGTGTACG	AGTGCCACTC	4300
CACGGATGCC	ACATCCATCT	TGGGCATCGG	CACTGTCTTT	GACCAAGCAG	4350
AGACTGCGGG	GGGAGACTG	GTTGTGCTCG	CCACTGCTAC	CCCTCCGGGC	4400
TCCGTCACTG	TGTCCCATCC	TAACATCGAG	GAGGTGTCTC	TGTCCACCAC	4450
CGGAGAGATC	CCCTTTTACG	GCAAGGCTAT	CCCCCTCGAG	GTGATCAAGG	4500
GGGAAGACA	TCTCATCTTC	TGCCACTCAA	AGAAGAAGTG	CGACGAGCTC	4550
GCCGGAAGC	TGGTCCGATT	GGGCATCAAT	GCCGTGGCCT	ACTACCGCGG	4600
TCTTGACGTG	TCTGTTCATC	CGACCGCGG	CGATGTTGTC	GTCGTGTCCA	4650
CCGATGCTCT	CATGACTGGC	TTTACCGCGG	ACTTCGACTC	TGTGATAGAC	4700
TGCAACACGT	GTGTCACTCA	GACAGTCGAT	TTGAGCCTTG	ACCTTACCTT	4750
TACCATTTGAG	ACAACCAAGC	TCCCCAGGA	TGCTGTCTCC	AGGACTCAAC	4800
GCCGGGGCAG	GACTGGCAGG	GGGAAGCCAG	GCATCTATAG	ATTGTGTCCA	4850
CCGGGGGAGC	GCCCCCTCCG	CATGTTCCAG	TGTTCCGTCC	TCTGTGAGTG	4900
CTATGACCGG	GGCTGTGCTT	GGTATGAGCT	CAGGCCCCGC	GAGACTACAG	4950
TTAGGCTACG	AGCGTACATG	AACACCCCGG	GGCTTCCCGT	GTTGCCAGGAC	5000
CATCTTGAAT	TTTGGGAGGG	CGTCTTTACG	GGCTCACCTC	ATATAGATGC	5050
CCACTTTTTA	TCCCAGACAA	AGCAGAGTGG	CGAGAACTTT	CCTTACCTGG	5100
TAGCGTACCA	AGCCACCGTG	TGCGCTAGGG	CTCAAGCCCC	TCCCCCATCG	5150
TGGGACCAGA	TGTGGAAGTG	TTTGATCCGC	CTTAAACCCA	CCCTCCATGG	5200
GCCAACACCC	CTGCTATACA	GACTGGGGGC	TGTTCCAGAAT	GAAGTCACCC	5250
TGACGCACCC	AATCACCAAA	TACATCATGA	CATGCATGTC	GGCCGACCTG	5300
GAGGTGCTCA	CGAGCACTG	GGTCTCGTT	GGGGGGTCC	TGGCTGCTCT	5350
GGCCCGCTAT	TGCTGTCAA	CAGGCTGGGT	GGTCATAGTG	GGCAGGATCG	5400
TCTTGTCCCG	GAAGCCGGCA	ATTATACCTG	ACAGGGAGGT	TCTCTACCAG	5450
GAGTTCGATG	AGATGGAAGA	GTTGCTCTCAG	CACTTACCGT	ACATCGAGCA	5500
AGGGATGATG	CTCGCTGAGC	AGTTCAAGCA	GAAGGCCCTC	GGCTCTCTGC	5550
AGACCGCGTC	CCGCCATGCA	GAGGTATCA	CCCCGTGCTG	CCAGACCAAC	5600
TGGCAGAAAC	TGAGGTCTT	TTGGGCGAAG	CACATGTGGA	ATTTCAATCAG	5650
TGGGATACAA	TACTTGGCGG	GCCTGTCAAC	GCTGCCCTGGT	AACCCCGCCA	5700

FIG. 4C

H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TTGCTTCATT	GATGGCTTTT	ACAGCTGCGG	TCACCAGCCC	ACTAACCCT	5750
GGCCAAACCC	TCTCTTCAA	CATATTGGGG	GGGTGGGIGG	CTGCCCAGCT	5800
CGCCGCCCCC	GGTCCCGCTA	CTGCCCTTTGT	GGGTGCTGGC	CTAGCTGGCG	5850
CCGCCATCGG	CAGCGTTGGA	CTGGGGAAGG	TCTGTGTGGA	CATTCTTGCA	5900
GGGTATGGCG	CGGGCGTGGC	GGGAGCTCTT	GTAGCATTC	AGATCATGAG	5950
CGGTGAGGTC	CCCTCCACGG	AGGAOCTGGT	CAATCTGCTG	CCCGCATCC	6000
TCTGGOCTGG	AGCCCTGTGA	GTGGGTGTGG	TCTGGCGAGC	AATACTGGCG	6050
CGGCAGTTG	GCCCGGGCGA	GGGGGCAGTG	CAATGGATGA	ACCGGCTAAT	6100
AGCCTTCGCC	TCCCGGGGGA	ACCATGTTTC	CCCCACGCAC	TACGTGCGGG	6150
AGAGCGATGC	AGCCGCCCGC	GTCAGTGCCA	TACTCAGCAG	CCTCACTGTA	6200
ACCCAGCTCC	TGAGGCGACT	GCATCAGTGG	ATAAGCTGG	AGTGTACCAC	6250
TCCATGCTCC	GGTTCTTGGC	TAAGGGACAT	CTGGGACTGG	ATATGCGAGG	6300
TGCTGAGCGA	CTTTAAGACC	TGGCTGAAAG	CCAAGCTCAT	GCCACAACCTG	6350
CCTGGGATTC	CCTTTGTGTC	CTGCCAGCGC	GGGTATAGGG	GGGTCTGGCG	6400
AGGAGACGGC	ATTATGCACA	CTGCTGCCA	CTGTGGAGCT	GAGATCACTG	6900
GACATGTCAA	AAACGGGACG	ATGAGGATCG	TGGTCCCTAG	GACCTGCAGG	6950
AACATGTGGA	GTGGGACGTT	CCCCATTAA	GCCTACACCA	CGGGCCCTTG	6550
TACTCCCTTT	CCTGGGCGGA	ACTATAAGTT	CGCGCTGTGG	AGGGTGTCTG	6600
CAGAGGAATA	CGTGGAGATA	AGGCGGGTGG	GGGACTTCCA	CTACGTATCG	6650
GGTATGACTA	CTGACAATCT	TAAATGCCCG	TGCCAGATCC	CATCGCCCGA	6700
ATTTTTTACA	GAATTGGACG	GGGTGCGCCT	ACACAGGTTT	GCGCCCCCTT	6750
GCAAGCCCTT	GCTGCGGGAG	GAGGTATCAT	TCAGAGTAGG	ACTCCACGAG	6800
TACCCGGTGG	GGTCCGAATT	ACCTTGCGAG	CCCGAACCGG	ACGTAGCCGT	6850
GTTGACGTCC	ATGCTCACTG	ATCCCTCCCA	TATAACAGCA	GAGGCGGCCG	6900
GGAGAAGGTT	GGCGAGAGGG	TCACCCCTTT	CTATGGCCAG	CTCCTGGGCT	6950
AGCCAGCTGT	CCGCTCCATC	TCTCAAGGCA	ACTTGCACCG	CCAACCATGA	7000
CTCCCTTGAC	GCCGAGCTCA	TAGAGGCTAA	CCTCCTGTGG	AGGCAGGAGA	7050
TGGGCGGCAA	CATCACCAGG	GTTGAGTCAG	AGAACAAAGT	GGTGATTCTG	7100
GACTCCTTCG	ATCCGCTTGT	GGCAGAGGAG	GATGAGCGGG	AGGTCCTCCGT	7150
ACCTGCAGAA	ATTCTGCCGA	AGTCTCGGAG	ATTGCCCCGG	GCCCTGCCCG	7200
TCTGGGCGCG	GCCGACTTAC	AACCCCCCGC	TAGTAGAGAC	GTGAAAAAG	7250
CCTGACTACG	AACCACCTGT	GGTCCATGGC	TGCCCGCTAC	CACCTCCACG	7300
GTCCCTCCTT	GTCCTCCCGC	CTCGGAAAAA	GCGTACGGTG	GTCCTCAACG	7350
AATCAACCTT	ATCTACTGCC	TTGGCCGAGC	TTGCCACCAA	AAGTTTTGGC	7400
AGCTCCTCAA	CTTCCGGCAT	TACGGGCGAC	AATACGACAA	CATCCTCTGA	7450
GCCCCCCCCC	TCTGGCTGCC	CCCCCGACTC	CGACGTTGAG	TCCTATTCTT	7500
CCATGCCCCC	CCTGGAGGGG	GAGCCTGGGG	ATCCGATCTT	CAGCGACGGG	7550
TCATGGTCCA	CGGTCACTAG	TGGGGCCGAC	ACGGAAGATG	TGTGTGTCTG	7600

FIG. 4D

H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CTCAATGTCT	TATTCCTGGA	CAGGCGCACT	CGTCACCCCG	TGCGCTGCGG	7650
AAGAACAAAA	ACTGCCCATC	AACGCACTGA	GCAACTCGTT	GCTAGCCCAT	7700
CACAATCTGG	TGTATTCCAC	CACCTCACGC	AGTGCTTGCC	AAAGGCAGAA	7750
GAAAGTCACA	TTTGACAGAC	TGCAAGTTCT	GGACAGCCAT	TACCAGGACG	7800
TGCTCAAGGA	GGTCAAAGCA	GCGGCGTCAA	AAGTGAAGGC	TAACTTGCTA	7850
TCCGTAGAGG	AAGCTTGCG	CCTGAGCCCG	CCACATTCAG	CCAAATCCAA	7900
GTTTGGCTAT	GGGGCAAAG	ACGTCCGTTG	CCATCCCGA	AAGGCCGTAG	7950
CCCACATCAA	CTCCGTGTGG	AAAGACCTTC	TGGAAGACAG	TGTAACAQCA	8000
ATAGACACTA	CCATCATGGC	CAAGAAGCAG	GTTTTCTGG	TTCAGCCTGA	8050
GAAGGGGGGT	CGTAAGCCAG	CTCGTCTCAT	CGTGTTCGCC	GACCTGGGCG	8100
TGCGCGTGTG	CGAGAAGATG	GCCCTGTACG	ACGTGGTTAG	CAAGCTCCCC	8150
CTGGCCGIGA	TGGGAAGCTC	CTACCGATTG	CAATACTCAC	CAGGACAGCG	8200
GGTTGAATTG	CTCGTGCAAG	CGTGGAAATC	CAAGAAGACC	CCGATGGGGT	8250
TCTCGTATGA	TACCCGCTGT	TTTGACTCCA	CAGTCACTGA	GAGCGACATC	8300
CGTACCGAGG	AGGCAATTTA	CCAATGTTGT	GACCTGGACC	CCCAAGCCCG	8350
CGTGGCCATC	AAGTCCCTCA	CTGAGAGGCT	TTATGTTGGG	GGCCCTCTTA	8400
CCAATTCAAG	GGGGGAAAC	TGCGGCTACC	GCAGGTGCCG	CGCGAGCGGC	8450
GTACTGACAA	CTAGCTGTGG	TAACACCCTC	ACTTGCTACA	TCAAGGCCCC	8500
GGCAGCCTGT	CGAGCCGCAG	GGCTCCAGGA	CTGCACCATG	CTCGTGTGTG	8550
GCGACGACTT	AGTCGTTATC	TGTGAAAGTG	CGGGGGTCCA	GGAGGACGCG	8600
GCGAGCCTGA	GAGCCTTCAC	GGAGGCTATG	ACCAGGTAAT	CGCCCCCCCC	8650
CGGGGACCCC	CCACAACCAG	AATACGACTT	GGAGCTTATA	ACATCATGCT	8700
CCTCCAACGT	GTCAGTCGCG	CACGACGGCG	CTGGAAAGAG	GGTCTACTAC	8750
CTTACCCGTG	ACCCATACAAC	CCCCCTCGCG	AGAGCCCGGT	GGGAGACAGC	8800
AAGACACACT	CCAGTCAATT	CCTGGCTAGG	CAACATAATC	ATGTTTGCCC	8850
CCACACTGTG	GGCGAGGATG	ATACTGATGA	CCCATTTCTT	TAGCGTCCTC	8900
ATAGCCAGGG	ATCAGCTTGA	ACAGGCTCTT	AACTGTGAGA	TCTACGGAGC	8950
CTGCTACTCC	ATAGAACCAC	TGGATCTACC	TCCAATCAAT	CAAAGACTCC	9000
ATGGCCCTCAG	CGCATTTTCA	CTCCACAGTT	ACTCTCCAGG	TGAAATCAAT	9050
AGGGTGGCCG	CATGCCTCAG	AAAACCTGGG	GTCCCGCCCT	TGCGAGCTTG	9100
GAGACACCGG	GCCCCGAGCG	TCCGCGCTAG	GCTTCTGTCC	AGAGGAGGCA	9150
GGGCTGCCAT	ATGTGGCAAG	TACCTCTTCA	ACTGGGCAGT	AAGAACAAAG	9200
CTCAAACCTCA	CTCCAATAGC	GGCCGCTGGC	CGGCTGGACT	TGTCCGGTTG	9250
GTTACCGGCT	GGCTACAGCG	GGGGAGACAT	TTATCACAGC	GTGTCTCATG	9300
CCCGGCCCCG	CTGGTCTCTG	TTTTTGCCAT	TCTGCTCGC	TGCAGGGGTA	9350
GGCATCTACC	TCCTCCCCAA	CCGATGAAGG	TTGGGGTAAA	CACTCCGGCC	9400
TCTTAAGCCA	TTTCTGTTT	TTTTTTTTTT	TTTTTTTTTT	TTTTTCTTTT	9450
TTTTTTTCTT	TCCTTTCCTT	CTTTTTTTTC	TTTCTTTTTT	CCTTCTTTAA	9500

FIG. 4E

H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TGGTGGCTCC	ATCTTAGCCC	TAGTCACGGC	TAGCTGTGAA	AGGTCCGTGA	9550
GCCGCATGAC	TGCAGAGAGT	GCTGATACTG	GCCTCTCTGC	AGATCATGT	9599

FIG. 4F

H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
MSINPKPQRK	TKRNINRRPQ	DVKFPGGGQI	VGGVYLLPRR	GPRLGVRATR	50
KTSESRQPRG	RRQPIPKARR	PEGRTWAQFG	YFWPLYGNEG	CGWAGWLLSP	100
RGSRPSWGPT	DPRRRSRNLG	KVIDILTCGF	ADLMGYIPLV	GAPLGGAARA	150
LAHGVRVLED	GVNYATGNLP	GCSFSIFLLA	LLSCLIVPAS	AYQVRNSSGL	200
YHVINDCPNS	SIVYEADAI	LHTPGCVPCV	REGNASRCW	AVTPTVATRD	250
GKLETTIQLRR	HIDLLVGSAT	LCSALYVGD	CGSVFLVGQL	FTFSPRRHWT	300
TQDCNCSIYP	GHITGHRMAW	IMMMNSPTA	ALVVAQLLRI	PQAIMIMTAG	350
AHWGVLAGIA	YFSMVGWAK	VLVLLLLFAG	VDAETHVTGG	NAGRITAGLV	400
GLLTPGAKQN	IQLININGSW	HINSTALN	ESLNTGWL	AG	450
GCPERLASCR	RLTDFAGWG	PISYANGSGL	DERPYCWHYP	PRPOGIVPAK	500
SVCGFVYCF	PSFVVGTID	RSGAPTYSWG	ANDIDVFVLN	NIRPPLGNWF	550
GCTWMNSTGF	TKVOGAPPCV	IGGVGNNTLL	CPIDCFRKH	P	600
FWITPRCMVD	YPYRLWHYPC	TINYTIFKVR	MYVGVVEHRL	EAACNWTRGE	650
RCDLEDNRDS	ELSPLLLSTT	QWQVLPCSFT	TLPALSTGLI	HLHQNTVDVQ	700
YLYGVGSSIA	SWAIKWEYV	LLFLLLADAR	VCSCLMMLL	ISQAEAALEN	750
LVILNAASLA	GIHGLVSFLV	FFCFAWYLKG	RWFGAVYAL	YGMWPLLLLL	800
LALPQRAYAL	DTEVAASCGG	VVLVGLMALT	LSPYYKRYIS	WCMWWLQYFL	850
TRVEAQLHW	VPPLNVRGGR	DAVILLMCV	HPILVFDITK	LLLAIFGPLW	900
ILQASLLKVP	YFVRVQGLR	ICALARKIAG	GHYVQMAI	IK	950
NHLLPLRDWA	HNGLRDLAVA	VEPVVFSRME	TKLITWGADT	AACGDIINGL	1000
PVSARRGQEI	LLGPADGMVS	KGWRLAPIT	AYAQQTRGLL	GCIITSLTGR	1050
DKNQVEGEVQ	IVSTATQTFL	ATCINGVCWT	VYHGAGIRTI	ASPKGPVTQM	1100
YTNVDQDLVG	WPAPQGSRL	TPCTCGSSDL	YLVTRHADVI	PVRRRGDSRG	1150
SLLSRPPISY	LKGSSGGPLL	CPAGHAVGLF	RAAVCTRGVA	KAVDFIPVEN	1200
LGTIMRSPVF	TINSSPPAVP	QSFQVAHLHA	PTGSGKSTKV	PAAYAAQGYK	1250
VLVLNPSVAA	TLGFGAYMSK	AHGVDENIRT	GVRTTTTGSP	ITYSTYGFEL	1300
ADGGCSGGAY	DIICDECHS	TDATSIILGIG	TVLDQAETAG	ARLWVLATAT	1350
PPGSVTVSHP	NIEEVALSTT	GEIPFYGKAI	PLEVIKGRH	LIFCHSKKKC	1400
DELAACKLVAL	GINAVAYYRG	LIVSVIPTSG	DVVVVSTDAL	MIGFTGDFDS	1450
VIDCNICVTQ	TVDFSLDPTF	TIETITLTPQD	AVSRTQRRGR	TGRGKPGIYR	1500
FVAPGERPSG	MFDSSVLCBC	YDAGCAWYEL	TPAETTIVRLR	AYMNTFGLPV	1550
QQDHLEFWEG	VFTGLTHIDA	HFLSQTKQSG	ENFPYLVAYQ	ATVCARAQAP	1600
PPSWDQMWKC	LIRLKPTLHG	PTPLLYRLGA	VQNEVTLTHP	ITKYIMICMS	1650
ADLEVVTSTW	VLVGGVLAAL	AAYCLSTGCV	VIVGRIVLSG	KPAIIPDREV	1700
LYQEFDEMEME	CSQHLPEYIEQ	GMMLAEQFKQ	KALGILLQTAS	RHAEVITPAV	1750
QTNWQKLEVF	WAKHMMNFIS	GIQYLAGLST	LPGNPAIASL	MAFTAAVTSP	1800
LTTGQTLLEN	ILGGWAAQL	AAPGAATAFV	GAGLAGAAIG	SVGLGKVLVD	1850
ILAGYGAGVA	GALVAFKIMS	GEVPSTEDLV	NLLPAILSEG	ALVGVWCAA	1900

FIG. 4G

H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ILRRHVGPGE	GAVQWMNRLI	AFASRGNHVS	PIHYVPESDA	AARVTAILSS	1950
LTVTQLLRRL	HQWISSECTT	PCSGSWLRDI	WDWICEVLSD	FKTWLKAKLM	2000
PQLPGIPFVS	CQRGYRGWR	GDGIMHIRCH	CGAETTGHVK	NGIMRIVGPR	2050
TCRNMWSGTF	PINAYTTGPC	TPLPAPNYKF	ALWRVSAEY	VEIRRVGDFH	2100
YVSGMTIDNL	KCPCQIPSPE	FFTELDGVRL	HRFAPPCKPL	LREEVSFRVG	2150
LHEYVPGSQL	PCEPEPDVAV	LTSMLIDPSH	ITAEAAGRRL	ARGSPPSMAS	2200
SSASQLSAPS	LKATCTANHD	SPDAELIEAN	LLWRQEMGGN	ITRVESENKV	2250
VILDSFDPLV	AEEDEREVS	PAETLRKSRR	FARALPWAR	PDYNPELVET	2300
WKKPDYEPFV	VHGCPPLPPR	SPPVPPPRKK	RIVVLTESTL	STALAEIATK	2350
SFGSSSTSGI	TGDNITTSSE	PAPSGCPPDS	DVESYSSMPP	LEGEFGDPDL	2400
SDGSWSTVSS	GADTEDVCC	SMSYSWTGAL	VTPCAAEEQK	LPTINALSNSL	2450
LRHHNLVYST	TSRSACQRQK	KVTFDRLQVL	DSHYQDVLKE	VKAAASKVKA	2500
NLLSVEEACS	LTPPHSAKSK	FGYGAKDVRC	HARKAVAHIN	SWKDLLEDS	2550
VTPIDTTIMA	KNEVFCVQPE	KGGRKPARLI	VFPDLGVRVC	EKMALYDMS	2600
KLPLAVMGSS	YGFQYSPGQR	VEFLVQAWKS	KKTPMGFSYD	TRCFDSTIVTE	2650
SDIRTEEAIV	QCCDLDPQAR	VAIKSLTERL	YVGGPLINSR	GENCGYRRCR	2700
ASGVLTTSCG	NILTCYIKAR	AACRAAGLQD	CTIMLVCGDDL	VVICESAGVQ	2750
EDAASLRAFT	EAMTRYSAAP	GDPPQPEYDL	ELITSCSSNV	SVAHDGACKR	2800
VYYLTRDPTT	PLARAAWETA	RHTPVNSWL	NIIMFAPTLW	ARMILMIHFF	2850
SVLIARDQLE	QALNCETYGA	CYSIEPLDLP	PIIQRLHGLS	AFSLHSYSPG	2900
EINRVAACLR	KLGVPLRAW	RHRARSVRAR	LLSRGGRAAI	CGKYLENNAV	2950
RTIKLKLTPIA	AAGRLDLGW	FTAGYSGGDI	YHSVSHARPR	WFWFCILLLA	3000
AGVGTYLLFN	R				3011

FIG. 4H

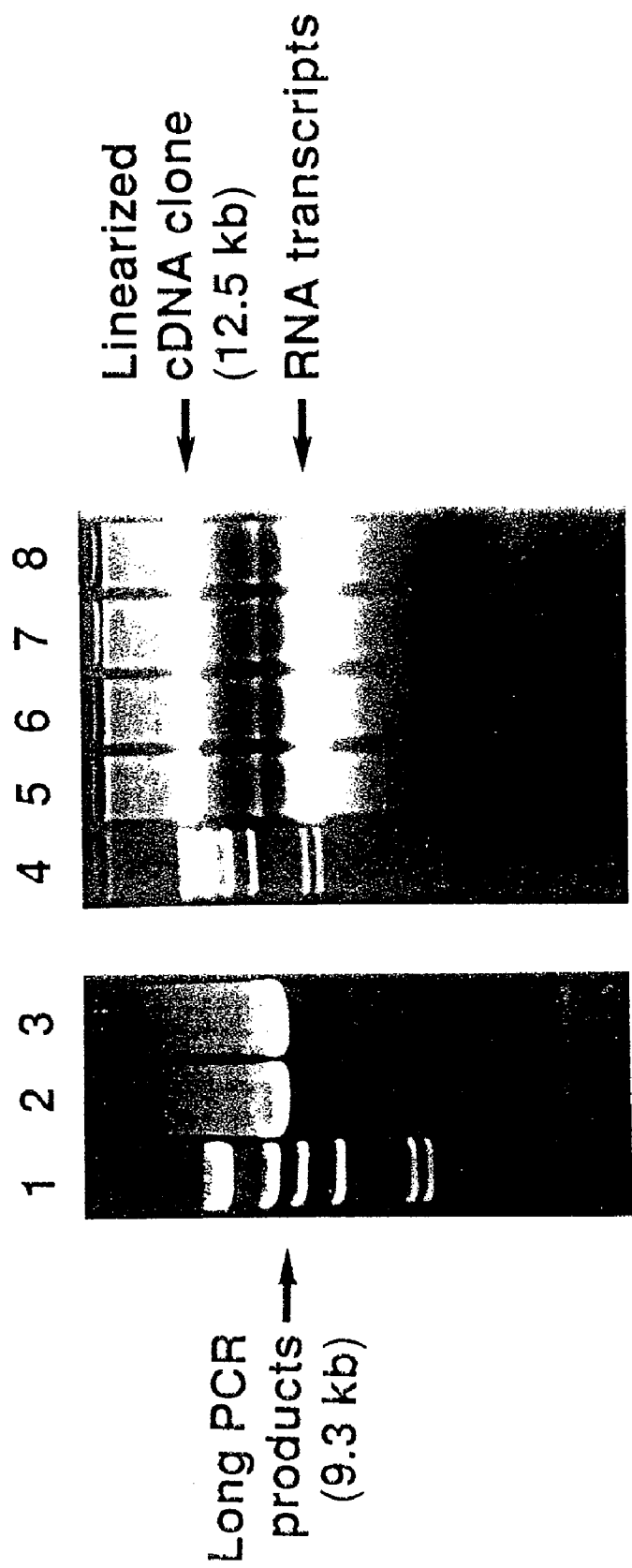


FIG. 5

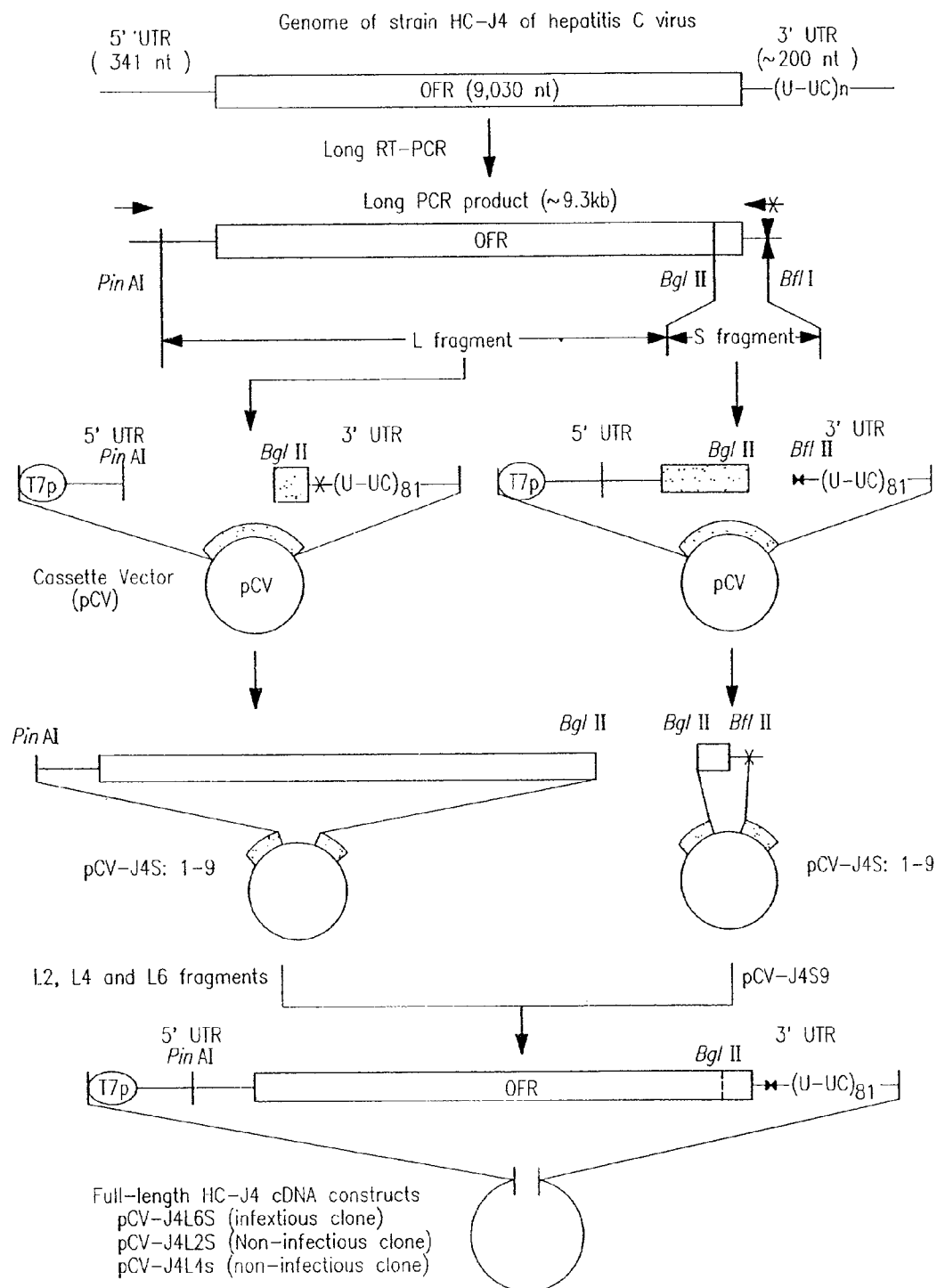


FIG. 6

	L fragment	Cons-p9	L1*(A)	L2(A)	L6(A)	L8(A)	L9(A)	L3(B)	L7*(B)	L10(B)	L4(C)	Cons-D	Cons-F
Core	16	N	S	.	.	.	.	.	.	.	.	.	N
	36	L	.	.	.	P	.	.	.	.	.	.	L
	52	A	.	.	.	.	.	T	T	T	T	T	A,T
	70	R	.	.	.	.	.	Q	Q	Q	.	R,Q	R,Q
	189	A	.	.	.	.	.	.	T	.	.	.	A
E1	195	R	.	.	.	.	.	H	.	H	.	.	R
	231	R	.	Q	Q	Q	.	.	.	.	.	.	R
	233	G	.	.	.	.	.	A	A	A	.	.	G
	234	N	.	.	.	.	.	D	D	D	.	.	N
	250	N	.	.	.	.	.	.	.	.	D	.	N
	299	E	A	.	.	.	A	.	.	.	.	.	E
	304	C	.	V	.	.	.	.	.	.	.	.	C
	379	A	.	.	.	.	.	T	.	T	.	.	A

FIG. 7A

	L fragment	Cons-p9	L1* (A)	L2(A)	L6(A)	L8(A)	L9(A)	L3(B)	L7* (B)	L10(B)	L4(C)	Cons-D	Cons-F
E2	384	E	.	.	.	.	.	T	T	T	.	E,T	E,T
	386	H	.	.	.	.	.	V	V	V	.	H,Y	H,Y
	388	T	.	.	.	.	.	S	S	S	.	T,S	T,S
	390	R	.	.	.	.	.	G	G	G	.	G	R,G
	391	V	.	.	.	.	.	.	.	A	.	.	V
	392	A	V	.	.	V	V	.	.	.	V	V	A,V
	394	H	.	.	.	.	.	R	R	R	R	.	H
	405	S	.	.	.	.	.	.	P	.	.	.	S
	434	Q	.	.	.	.	.	H	H	H	.	H	Q,H
	438	F	.	.	.	.	.	L	L	L	L	L	F,L
	444	A	.	.	.	.	.	T	T	T	T	T	A,T
	450	S	.	.	.	.	.	.	.	.	P	.	S
	458	S	.	.	.	N	.	.	.	.	.	.	S
	466	A	.	.	.	.	.	V	V	V	.	A,V	A,V
	474	Y	.	.	.	.	.	H	.	.	.	.	Y
	476	K	.	.	.	.	.	E	E	E	E	E	K,E
	496	V	.	.	.	.	.	I	I	I	I	I	V,I
	524	V	.	.	.	.	A	.	A	.	.	.	V
	536	V	.	M	.	.	.	.	.	.	.	.	V
	580	I	.	.	.	.	.	.	V	.	.	.	I
	622	L	V	.	.	.	V	.	.	.	.	.	L
	673	Q	.	.	.	P	.	.	.	.	.	.	Q
	783	A	.	.	.	.	.	.	V	.	.	.	A
p7													

FIG. 7B

	L fragment	Cons-p9	L1*(A)	L2(A)	L6(A)	L8(A)	L9(A)	L3(B)	L7*(B)	L10(B)	L4(C)	Cons-D	Cons-F
NS2	820	G	.	.	.	.	.	.	S	.	.	.	G
	857	M	.	.	.	.	.	I	.	.	.	.	M
	927	K	.	.	.	.	.	.	R	.	.	.	K
	934	V	I	I	.	I	I	.	.	.	.	.	V
	937	A	.	.	V	.	.	.	.	.	.	.	A
	978	A	.	.	.	.	.	D	D	D	.	D	A,D
	1028	P	.	.	.	S	.	.	.	.	.	.	P
NS3	1031	A	.	.	.	.	.	.	.	T	.	.	A
	1043	V	.	.	I	.	.	I	I	I	.	I	V,I
	1067	Q	.	.	.	.	.	H	H	H	.	Q,H	Q,H
	1097	I	X	.	.	.	.	.	.	.	.	.	I
	1188	G	R	.	.	.	.	.	.	.	.	.	G
	1215	S	.	.	T	.	.	.	.	.	.	.	S
	1223	F	.	S	.	.	.	.	.	.	.	.	F
	1226	A	.	.	.	.	.	.	.	V	.	.	A
	1339	A	V	.	.	.	.	.	.	.	.	.	A
	1399	K	N	.	.	.	.	.	.	.	.	.	K
	1503	T	.	.	.	.	.	S	.	S	.	.	T
	1528	Y	.	.	.	.	.	.	.	.	H	.	Y
	1535	T	A	.	.	.	.	.	.	.	.	.	T
NS4A	1662	L	.	P	.	.	.	.	.	.	.	.	L

FIG. 7C

	L fragment	Cons-p9	L1*(A)	L2(A)	L6(A)	L8(A)	L9(A)	L3(B)	L7*(B)	L10(B)	L4(C)	Cons-D	Cons-F
NS4B	1753	K	•	R	•	•	•	•	•	•	•	•	K
	1805	H	•	•	N	•	•	N	•	N	N	N	H,N
	1949	S	•	•	•	•	•	•	•	•	P	•	S
	2105	M	•	•	•	•	V	I	•	I	•	•	M
	2136	K	•	•	•	•	•	•	•	•	R	•	K
NS5A	2146	T	•	•	•	•	•	A	A	A	•	I,A	T,A
	2226	L	•	•	•	•	•	P	•	•	•	•	L
	2259	L	•	•	•	•	•	F	•	•	•	•	L
	2262	E	•	•	•	•	•	D	D	D	•	E,D	E,D
	2334	V	•	•	•	•	•	I	•	•	•	•	V
	2371	L	•	•	•	•	•	Q	Q	Q	•	L,Q	L,Q
	2385	Y	•	•	•	•	•	•	•	•	H	•	Y
	2692	N	•	•	•	•	•	•	S	•	•	•	N
	2757	A	•	•	•	•	•	•	•	•	•	•	A
	2785	C	•	R	•	•	•	•	•	•	•	•	C
NS5B	2824	I	•	V	•	•	•	•	•	•	•	•	I
	2861	A	•	•	•	•	•	V	•	•	•	•	A
	S fragment		S5	S9	S2	S3	S7	S8	S10	S4	S6		
	2968	G	•	•	•	•	•	S	S	•	•	•	G
	2975	S	•	•	•	•	•	G	G	G	G	•	S
	2978	D	•	•	•	•	•	•	•	•	G	•	D
	2999	S	•	F	F	F	•	•	•	•	•	•	S

FIG. 7D

aa	nt	L1 (A)	L2 (A)	L6 (A)	L8 (A)	L9 (A)	L3 (B)	L7 (B)	L10 (B)	L4 (C)	HC-J4/91	HC-J4/83
L1 (A)			0.56	0.60	0.36	0.33	1.50	1.53	1.46	0.95	0.83	1.79
L2 (A)	0.59			0.55	0.35	0.50	1.49	1.51	1.45	0.98	0.82	1.77
L6 (A)	0.52	0.42			0.31	0.55	1.33	1.38	1.29	0.80	0.68	1.58
L8 (A)	0.42	0.38	0.31			0.31	1.32	1.34	1.28	0.79	0.65	1.62
L9 (A)	0.35	0.52	0.45	0.35			1.42	1.42	1.38	0.91	0.75	1.66
L3 (B)	1.47	1.43	1.15	1.33	1.22	1.36		0.61	0.30	1.43	0.90	1.51
L7 (B)	1.36	1.33	1.05	1.22	1.22	1.22	0.66		0.57	1.47	0.95	1.54
L10 (B)	1.36	1.33	0.59	1.22	1.22	1.26	0.31	0.56		1.37	0.85	1.42
L4 (C)	0.77	0.80	0.59	0.63	0.63	1.26	1.12	1.08	1.01		0.76	1.73
HC-J4/91	0.94	0.91	0.63	0.80	0.80	0.87	0.77	0.73	0.66	0.52		1.22
HC-J4/83	1.96	1.89	1.68	1.85	1.85	1.82	1.75	1.61	1.61	1.71	1.40	

FIG. 8

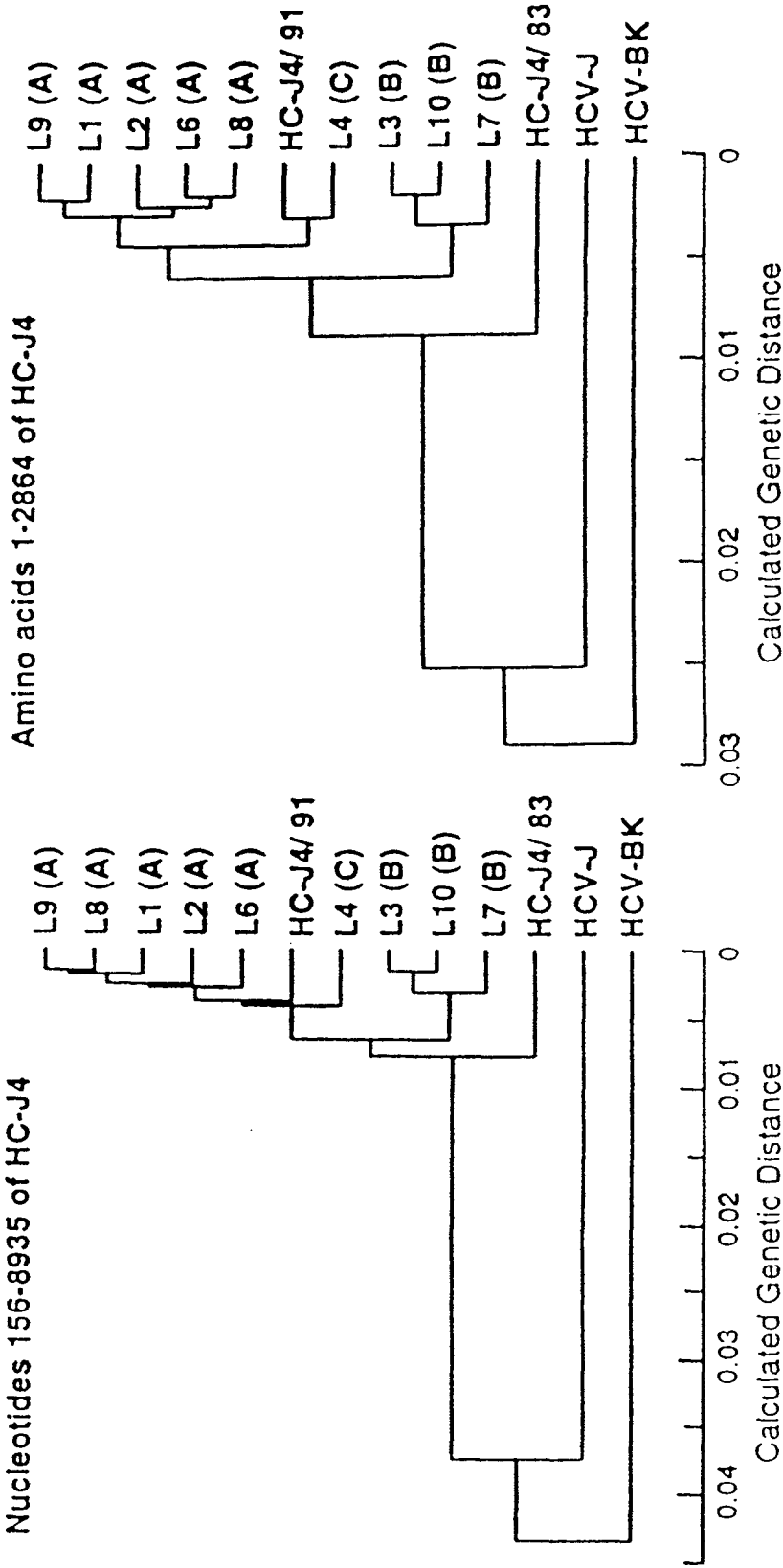


FIG. 9

	379	413	468	486
HC-J4L6 (A)	: AGVDG	ETHTGRVAGHTTSGFTSLESSGAS QKIQL	GWGPIT	YTKPNSS DQRPYC
HC-J4L2 (A)	:			
HC-J4/91-20	:	 R		 E
HC-J4L1 (A)	:	 V		
HC-J4L8 (A)	:	 V		
HC-J4L9 (A)	:	 V		
HC-J4/91-21	:	 V		 G
HC-J4L4 (C)	:	 V.R		 E
HC-J4/91-23	:	 V.R		 E
HC-J4/91-22	:	 V.R	 A	 E
HC-J4L7 (B)	:	 T.Y.S.G P		 E
HC-J4L10 (B)	: T	 T.Y.S.GA		 E
HC-J4L3 (B)	: T	 T.Y.S.G		 H.E
HC-J4/91-26	: T	 T.Y.S.G		 G.D.L
HC-J4/91-25	:	 A.Y.S.G		 E
HC-J4/91-24	:	 A.Y.S.G		 E P
HC-J4/91	:	 A.Y.S.G		 E P
HC-J4/91-27	:	 K.Y.S.GA.S P R		 ESG.R
HC-J4/83	:	 Y.S.GA.S TLA P R		 E.D.P

↕
HVR2

↕
HVR1

FIG. 10

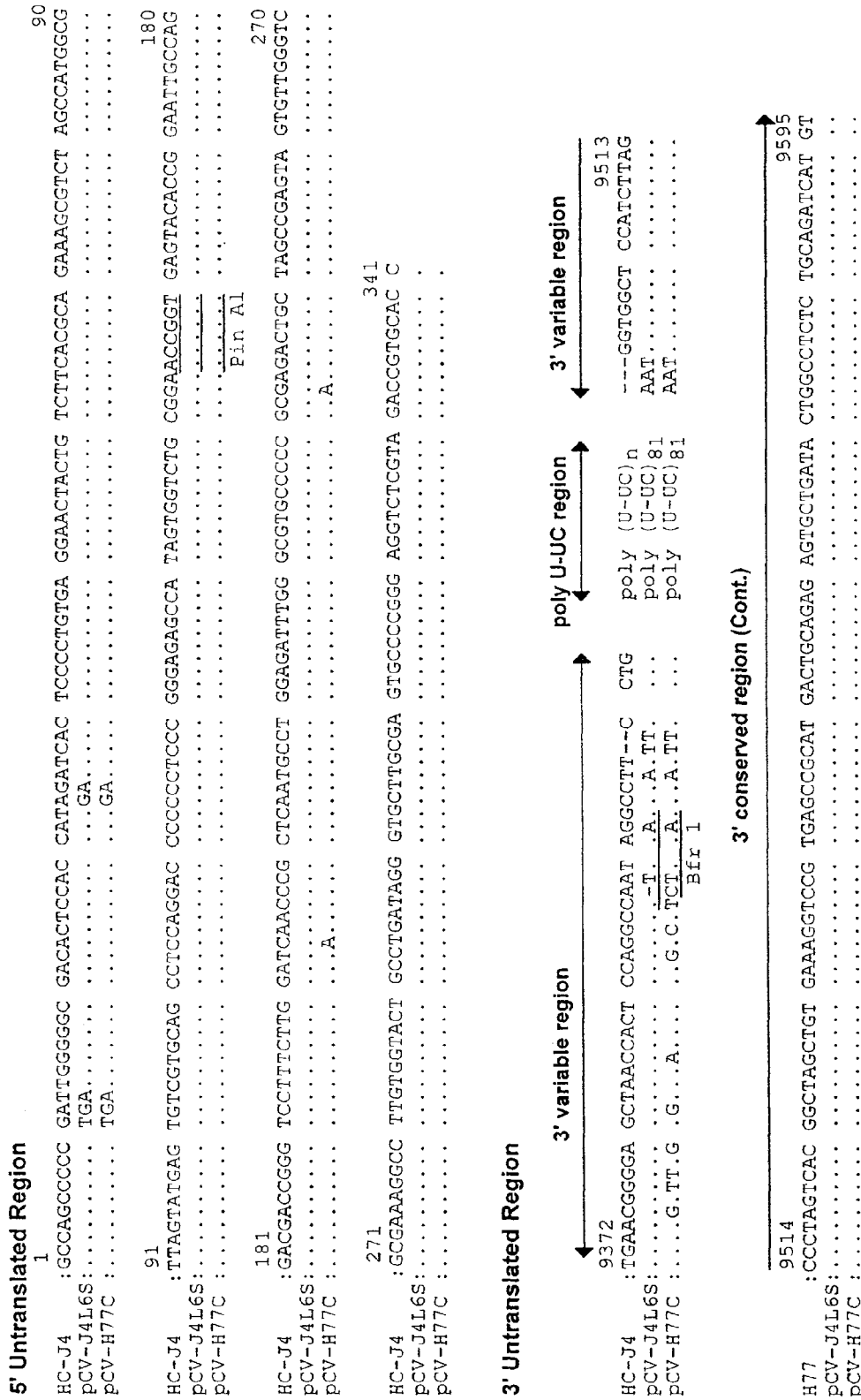


FIG. 11

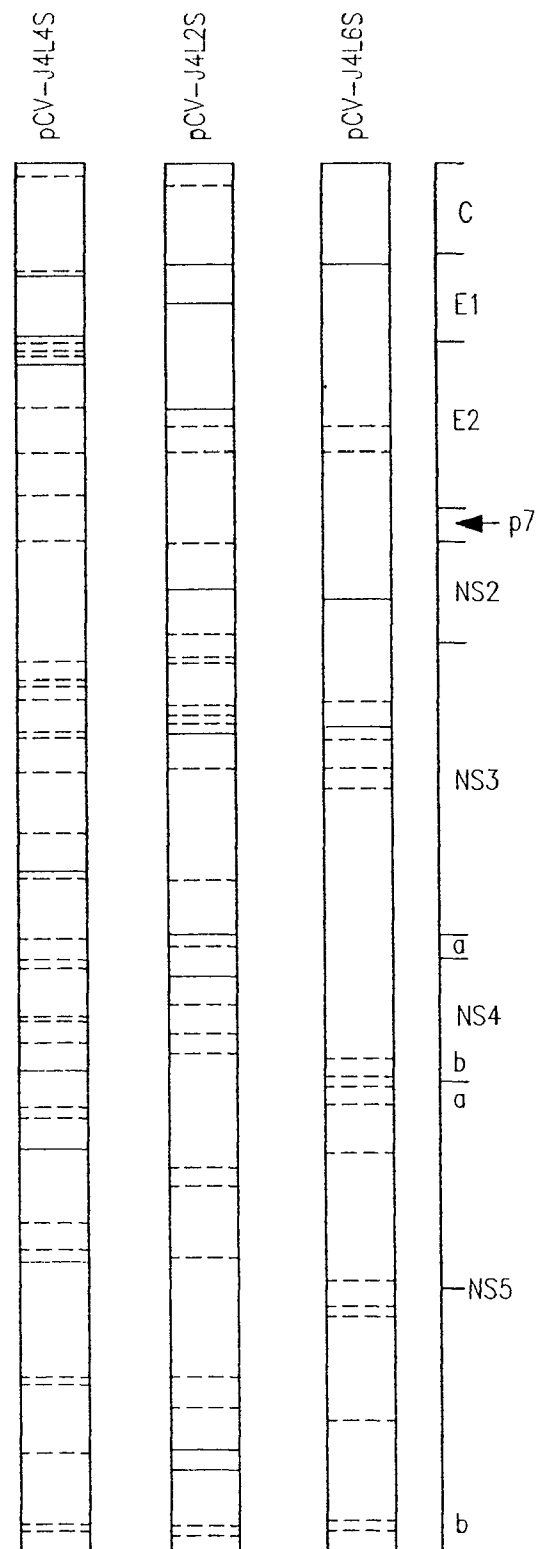


FIG. 12

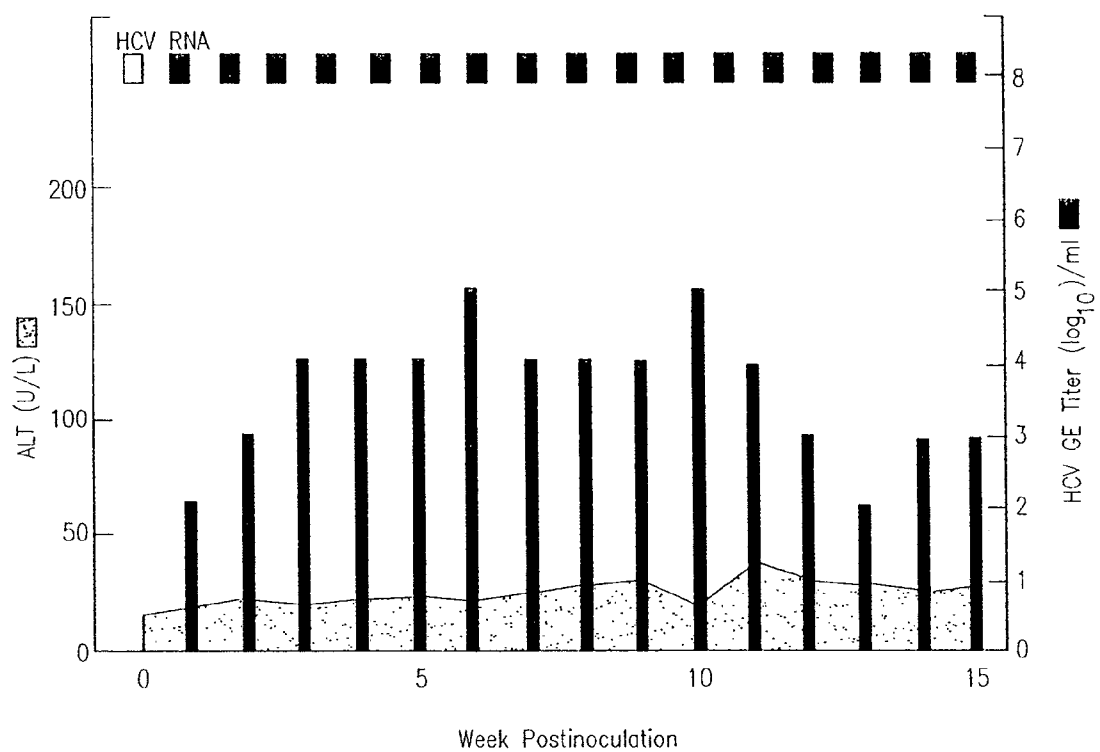


FIG. 13

HC-J4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GCCAGCCCC	TGATGGGGC	GACACTCCAC	CATGAATCAC	TCCCCTGTGA	50
GGAACTACTG	TCTTCACGCA	GAAAGCGTCT	AGCCATGGCG	TTAGIATGAG	100
TGTGCTGCAG	CCTCCAGGAC	CCCCCCTCCC	GGGAGAGCCA	TAGTGGTCTG	150
CGGAACCGGT	GAGTACACCG	GAATTGCCAG	GACGACCGGG	TCCTTTCTTG	200
GATCAACCCG	CTCAATGCCT	GGAGATTTGG	GCGTGCCCC	GCGAGACTGC	250
TAGCCGAGTA	GTGTTCGGTC	GCGAAAGGCC	TTGTGGTACT	GCCTGATAGG	300
GTGCTTGCGA	GTGCCCCGGG	AGGTCTCGTA	GACCGTGCAC	CATGAGCACG	350
AATCCTAAAC	CTCAAAGAAA	AACCAAACGT	AACACCAACC	GCCGCCACA	400
GGACGTCAAG	TTCCCCGGCG	GTGGTCAGAT	CGTTGGTGA	GTTTACCTGT	450
TGCCCCGAG	GGGCCCCAGG	TTGGGTGTGC	GCGCGACTAG	GAAGGCTTCC	500
GAGCGGTGCG	AACCTCGTGG	AAGGCGACAA	CCTATCCCAA	AGGCTCGCCG	550
ACCCGAGGGC	AGGGCCTGGG	CTCAGCCCCG	GTACCCCTTG	CCCCCTCTATG	600
GCAATGAGGG	CCTGCGGTGG	GCAGGATGGC	TCCTGTCAAC	CCGCGGCTCC	650
CGGCCTAGTT	GGGCCCCCAC	GGACCCCCGG	CGTAGGTGCG	GTAACCTGGG	700
TAAGGTCATC	GATACCCCTA	CATGCGGCTT	CGCCGATCTC	ATGGGGTACA	750
TTCCGCTCGT	CGGCGCCCCC	CTAGGGGGCG	CTGCCAGGGC	CTTGGCACAC	800
GGTGTCCGGG	TTCTGGAGGA	CGGCGTGAAC	TATGCAACAG	GGAACCTTGC	850
CGGTTGCTCT	TTCTCTATCT	TCCTCTTGCC	TCTGCTGTCC	TGTTTGACCA	900
TCCCAGCTTC	CGCTTATGAA	GTGCGCAACG	TGTCCGGGAT	ATACCATGTC	950
ACGAACGACT	GCTCCAACCT	AAGCATTTGT	TATGAGGCAG	CGGACGTGAT	1000
CATGCATACT	CCCGGGTGGG	TGCCCCGTGT	TCAGGAGGGT	AACAGCTCCC	1050
GTTGCTGGGT	AGCGCTCACT	CCCACGCTCG	CGGCCAGGAA	TGCCAGCGTC	1100
CCCACTACGA	CAATACGACG	CCACGTGCAC	TTGCTCGTTG	GGACGGCTGC	1150
TTTCTGCTCC	GCTATGTACG	TGGGGGATCT	CTGCCGATCT	ATTTTCCTCG	1200
TCTCCCAGCT	GTTCACCTTC	TGCCCCGCGC	GGCATGAGAC	AGTGCAGGAC	1250
TGCAACTGCT	CAATCTATCC	CGGCCATGTA	TCAGGTCAAC	GCATGGCTTG	1300
GGATATGATG	ATGAACTGGT	CACCTACAAC	AGCCCTAGTG	GTGTGCGAGT	1350
TGCTCCGGAT	CCCACAAGCT	GTCGTGGACA	TGGTGGCGGG	GGCCCACTGG	1400
GGAGTCCTGG	CGGGCCCTGC	CTACTATTCC	ATGGTAGGGA	ACTGGGCTAA	1450
GGTTCTGATT	GTGGCGCTAC	TCTTTGCGCG	CGTTGACGGG	GAGACCCACA	1500
CGACGGGGAG	GGTGGCCGGC	CACACCACCT	COGGGTTCAC	GTCCCTTTTC	1550
TCATCTGGGG	CGTCTCAGAA	AATCCAGCTT	GTGAATACCA	ACGGCAGCTG	1600
GCACATCAAC	AGGACTGCCC	TAAATTGCAA	TGACTCCCTC	CAAACCTGGT	1650
TCTTTGCGCG	GCTGTTTAC	GCACACAAGT	TCAACTCGTC	CGGGTGCCCC	1700
GAGCGCATGG	CCAGCTGCCC	CCCCATTGAC	TGGTTCGCCC	AGGGGTGGGG	1750
CCCCATCACC	TATACTAAGC	CTAACAGCTC	GGATCAGAGG	CCTTATTGCT	1800
GGCATTACGC	GCTCGAACCG	TGIGGTGTGC	TACCCGCGTC	GCAGGTGTGT	1850
GGTCCAGTGT	ATTGTTTCAC	CCCAAGCCCT	GTTGTGGTGG	GGACCACCGA	1900

FIG. 14A

HC-J4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TGCTTCCGGT	GTCCCTACGT	ATAGCTGGGG	GGAGAATGAG	ACAGACGTGA	1950
TGCTCCTCAA	CAACACGGGT	CCGCCACAAG	GCAACTGGTT	CGGCTGTACA	2000
TGGATGAATA	GTACTGGGTT	CACTAAGACG	TGCGGAGGTC	CCCCGTGTAA	2050
CATCGGGGGG	GTCCGTAAAC	GCACCTTGAT	CTGCCCCACG	GA CTGCTTCC	2100
GGAAGCACCC	CGAGGCTACT	TACACAAAAT	GTGGCTCGGG	GCCCTGGTTG	2150
ACACCTAGGT	GCCTAGTAGA	CTACCCATAC	AGGCTTTGGC	ACTACCCCTG	2200
CACTCTCAAT	TTTTCCATCT	TTAAGGTTAG	GATGTATGTG	GGGGGCGTGG	2250
AGCACAGGCT	CAATGCCGCA	TGCAATTGGA	CTCGAGGAGA	GCGCTGTAAAC	2300
TTGGAGGACA	GGGATAGGTC	AGAACTCAGC	CCGCTGCTGC	TGTCTACAAC	2350
AGAGTGGCAG	ATACTGCCCT	GTGCTTTTCAC	CACCCCTACCG	GCTTTTATCCA	2400
CTGGTTTIGAT	CCATCTCCAT	CAGAACATCG	TGGACGTGCA	ATACCTGTAC	2450
GGTGTAGGGT	CAGCGTTTGT	CTCCTTTTCCA	ATCAAATGGG	AGTACATCCT	2500
GTTCCTTTTC	CTTCTCCTGG	CAGACGCGCG	CGTGTGTGCC	TGCTTGTGGA	2550
TGATGCTGCT	GATAGCCCGAG	GCTGAGGCGG	CCTTAGAGAA	CTTGGTGGTC	2600
CTCAATGCCG	CGTCCGTGGC	CGGAGCGCAT	GGTATTCTCT	CCTTTCTTGT	2650
GTTCCTTCTG	GCGCCTGGT	ACATTAAAGG	CAGGCTGGCT	CCTGGGGCGG	2700
CGTATGCITT	TTATGGCGTA	TGGCCGCTGC	TCCTGCTCCT	ACTGGCGTTA	2750
CCACCACGAG	CTTACGCCTT	GGACCGGGAG	ATGGCTGCAT	CGTCCGGGGG	2800
TGCGGTCTTT	GTAGGTCTGG	TATTCTTGAC	CTTGTACCA	TACTACAAAG	2850
TGTTTCTCAC	TAGGCTCATA	TGGTGGTTAC	AATACTTTAT	CACCAGAGCC	2900
GAGGCGCACA	TGCAAGTGIG	GGTCCCCCCC	CTCAACGTTT	GGGAGGCGCG	2950
CGATGCCATC	ATCCTCCTCA	CGTGTGCGGT	TCATCCAGAG	TTAATTTTIG	3000
ACATCACCAA	ACTCCTGCTC	GCCATACTCG	GCCCCGCTCAT	GGTGTCCAG	3050
GCTGGCATAA	CGAGAGTGCC	GTACTTCGTG	CGCGCTCAAG	GGCTCATTCG	3100
TGCATGCATG	TTAGTGCGAA	AAGTCGCCGG	GGTCAATTAT	GTCCAAATGG	3150
TCTTCATGAA	GCTGGGCGCG	CTGACAGGTA	CGTACGTTTA	TAACCATCTT	3200
ACCCCACTGC	GGGACTGGGC	CCACGCGGGC	CTACGAGACC	TTGCGGTGGC	3250
GGTAGAGCCC	GTGCTCTTCT	CCGCCATGGA	GACCAAGGTC	ATCACCTGGG	3300
GAGCAGACAC	CGCTGCGTGT	GGGGACATCA	TCTTGGGTCT	ACCCGCTCTC	3350
GCCCCAAGGG	GGAAGGAGAT	ATTTTTTGGGA	CCGGCTGATA	GTCTCGAAGG	3400
GCAAGGGTGG	CGACTCCTTG	CGCCCATCAC	GGCCTACTCC	CAACAAACGC	3450
GGGGCGTACT	TGGTTGCATC	ATCACTAGCC	TCACAGGCGG	GGACAAGAAC	3500
CAGGTGGAAG	GGGAGGTTCA	AGTGGTTTCT	ACCGCAACAC	AATCTTTTCT	3550
GGCGACCTGC	ATCAACGGCG	TGTGCTGGAC	TGTCTACCAT	GGCGCTGGCT	3600
CGAAGACCC	AGCCGCTCCA	AAAGGTCCAA	TCACCCAAAT	GTACAACAAT	3650
GTAGACCTGG	ACCTCGTCGG	CTGGCAGGCG	CCCCCGGGG	CGCGCTCCAT	3700
GACACCATGC	AGCTGTGGCA	GCTCGGACCT	TTACTTGGTC	ACGAGACATG	3750
CTGATGTCAT	TCCGGTGGCG	CGGCGAGGCG	ACAGCAGGGG	AAGTCTACTC	3800

FIG. 14B

HC-J4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TC0000CAGGC	CCGTCTCCIA	CCTGAAAGGC	TCCTCGGGTG	GTCCATTGCT	3850
TTGCCCTTCG	GGGCACGTG	TGGGCGTCTT	CCGGGCTGCT	GTGTGCACCC	3900
GGGGGGTGGC	GAAGGCGGTG	GACTTCATAC	CCGTTGAGTC	TATGGAAACT	3950
ACCATGCGGT	CTCCGGTCTT	CACAGACAAC	TCAAC00000	CGGCTGTACC	4000
GCAGACATTG	CAAGTGGCAC	ATCTGCACGC	TCCTACTGGC	AGCGGCAAGA	4050
GCACCAAAGT	GCCGGCTGG	TATGCAGCCC	AAGGGTACAA	GGTGCTCGTC	4100
CTGAACCCGT	CCGTTGCCGC	CACCTTAGGG	TTTGGGGCGT	ATATGTCCAA	4150
GGCACACGGT	ATCGACCCIA	ACATCAGAAC	TGGGGTAAGG	ACCATTACCA	4200
CGGGCGGCTC	CATTACGTAC	TCCACCTATG	GCAAGTT0CT	TG00GACGGT	4250
GGCTGTCTCG	GGGGCGCCIA	TGACATCATA	ATATGTGATG	AGTGCCACTC	4300
AACTGACTCG	ACTA0CATCT	TGGGCATCGG	CACAGT0CTG	GACCAAGCGG	4350
AGACGGCTGG	AGCGCGGCTC	GTCGIGCTCG	CCACCGCTAC	ACCTCCGGGA	4400
TCGGTTACCG	TGCCACACCC	CAATATCGAG	GAAATAGG0C	TGTCCAACAA	4450
TGGAGAGATC	CCCTTCTATG	GCAAAGCCAT	CCCCATTGAG	GCCATCAAGG	4500
GGGGGAGGCA	TCTCATTTTC	TGCCATTCCA	AGAAGAAATG	TGACGAGCTC	4550
GCCGCAAAGC	TGACAGGCCCT	CGGACTGAAC	GCTGTAGCAT	ATTACCGGGG	4600
CCTTGATGTG	TCCGTCATAC	CGCCTATCGG	AGACGT0GTT	GTGTTGGCAA	4650
CAGACGCTCT	AATGACGGGT	TTCACCGCGG	ATTTTGACTC	AGTGATCGAC	4700
TGCAATACAT	GTGTACCCCA	GACAGTCGAC	TTCAGCTTGG	ATCCACCTTT	4750
CACCATTGAG	ACGACGACCG	TGCCCCAAGA	CGCGGTGTG	CGCTCGCAAC	4800
GGCGAGGTAG	AACTGGCAGG	GGTAGGAGTG	GCATCTACAG	GTTTGTGACT	4850
CCAGGAGAAC	GGCCCTCGGG	CATGTT0GAT	TCTTCGGTCC	TGTGTGAGTG	4900
CTATGACCGG	GGCTGTGCTT	GGTATGAGCT	CACG000CGT	GAGACCTCGG	4950
TTAGGTTGCG	GGCTTACCTA	AATACACCAG	GGTTGCCCGT	CTGCCAGGAC	5000
CATCTGGAGT	TCTGGGAGAG	CGTCTTCACA	GGCCTCACCC	ACATAGATGC	5050
CCACTTCCTG	TCCCAGACTA	AACAGGCAGG	AGACAACTTT	CCTTACCTGG	5100
TGGCATATCA	AGCTACAGTG	TGCGCCAGGG	CTCAAGCTCC	ACCTCCATCG	5150
TGGGACCAAA	TGTGGAAGTG	TCTCATACGG	CTGAAACCTA	CACTGCACGG	5200
GCCAACACCC	CTGCTGTATA	GGCTAGGAGC	CGTCCAAAAT	GAGGTTCATCC	5250
TCACACACCC	CATAACTAAA	TACATCATGG	CATGCATGTC	GGCTGACCTG	5300
GAGGTGCTCA	CTAGCA0CTG	GGTGCTGGTA	GGCGGAGTCC	TTGCAGCTTT	5350
GGCCGCATAC	TG0CTGACGA	CAGGCAGTGT	GGTCATTGTG	GGCAGGATCA	5400
TCTTGTCCCG	GAAGCCAGCT	GTCGTTCCCG	ACAGGGAAGT	CCTCTACCAG	5450
GAGTT0GATG	AGATGGAAGA	GTGTGCCTCA	CAACTT0CTT	ACATCGAGCA	5500
GGGAATGCAG	CTCGCCGAGC	AATTCAAGCA	AAAGGCGCTC	GGTGTGTGTC	5550
AAACGGCCAC	CAAGCAAGCG	GAGGCTGCTG	CTCCCGTGGT	GGAGTCCAAG	5600
TGGCGAGCCC	TTGAGACCTT	CTGGGCGAAG	CACATGTGGA	ATTTCATCAG	5650
CGGAATACAG	TACCTAGCAG	GCTTATCCAC	TCTG0CTGGA	AACCCCGCGA	5700

FIG. 14C

HC-J4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TAGCATCATT	GATGGCATT	ACAGCTTCTA	TCACTAGCCC	GCTCACCACC	5750
CAAAACACCC	TCCTGTTTAA	CATCTTGGGG	GGATGGGTGG	CTGCCCCAACT	5800
CGCTCCTCCC	AGCGCTGCGT	CAGCTTTTCGT	GGGCGCCGGC	ATCGCCGGAG	5850
CGGCTGTTGG	CAGCATAGGC	CTTGGGAAGG	TGCTOGTGA	CATCTTGGCG	5900
GGCTATGSGG	CAGGGGTAGC	CGGCGCACTC	GTGGCCTTTA	AGGTCATGAG	5950
CGGCGAGGIG	CCCTCCACCG	AGGACCTGGT	CAACTTACTC	CCTGCCATCC	6000
TCTCTCCTGG	TGCCCTGGTC	GTCGGGGTGG	TGTGCGCAGC	AATACTGGGT	6050
CGGCACGTGG	GCCCGGGAGA	GGGGGCTGTG	CAGTGGATGA	ACCGGCTGAT	6100
AGCGTTTCGT	TGCGGGGGTA	ACCACTCTCT	CCCTACGCAC	TATGTGCTTG	6150
AGAGCGACGC	TGCAGCAAGT	GTCACCTAGA	TCCTCTCTAG	CCTTACCATC	6200
ACTCAACTGC	TGAAGCGGCT	CCACCAGTGG	ATTAAATGAGG	ACTGCTCTAC	6250
GCCATGCTCC	GGCTCGTGGC	TAAGGGATGT	TTGGGATTGG	ATATGCAACG	6300
TGTTGACTGA	CTTCAAGACC	TGGCTCCAGT	CCAACTCCT	GCCGCGGTTA	6350
CCGGGAGTCC	CTTTCCCTGTC	ATGCCAACGC	GGGTACAAGG	GAGTCTGGCG	6400
GGGGGACGGC	ATCATGCAAA	CCACCTGCCC	ATGCGGAGCA	CAGATCGCCG	6450
GACATGTCAA	AAACGGTTCC	ATGAGGATCG	TAGGGCCTAG	AACCTGCAGC	6500
AACACGTGGC	ACGGAAAGTT	CCCCATCAAC	GCATACACCA	CGGGACCTTG	6550
CACACCCCTC	CCGGCGCCCA	ACTATTCCAG	GGCGCTATGG	CGGGTGGCTG	6600
CTGAGGAGTA	CGTGGAGGTT	ACGCGTGTGG	GGGATTTCOA	CTACGTGACG	6650
GGCATGACCA	CTGACCAAGT	AAAGTGCCCA	TGCCAGGTTT	CGGCCCCCGA	6700
ATTCTTCAAG	GAGGTGGATG	GAGTGGGGTT	GCACAGGTAC	GCTCCGGCGT	6750
GCAAACCTCT	TCTACGGGAG	GACGTCAAGT	TCCAGGTCCG	GCTCAACCAA	6800
TACTTGGTTC	GGTCGCAGCT	CCCATGCGAG	CCCGAACCGG	ACGTAACAGT	6850
GCTTACTTCC	ATGCTCACC	ATCCCTCCCA	CATTACAGCA	GAGACGGCTA	6900
AGCGTAGGCT	GGCTAGAGGG	TCTCCCCCTT	CTTTAGCCAG	CTCATCAGCT	6950
AGCCAGTTGT	CTGGCCCTTC	TTTGAAGGGG	ACATGCACCTA	CCCACCATGA	7000
CTCCCCGGAC	GCTGACCTCA	TGAGGGCCAA	CCTCTTGTGG	CGGCAGGAGA	7050
TGGGCGGAAA	CATCACTCGC	GTGGAGTCAG	AGAATAAGGT	AGTAATTCTG	7100
GACTCTTTTC	AACCGCTTCA	CGCGGAGGGG	GATGAGAGGG	AGATATCCGT	7150
CGCGGCGGAG	ATCCTGCGAA	AATCCAGGAA	GTTCCTCTCA	GCGTTGCCCA	7200
TATGGGCACG	CCCGGACTAC	AATCCTCCAC	TGCTAGAGTC	CTGGAAGGAC	7250
CCCGACTACG	TCCCTCCGGT	GGTACACGGA	TGCCCATTTG	CACCTACCAA	7300
GGCTCCTCCA	ATACCACCTC	CACGGAGAAA	GAGGACGGTT	GTCCTGACAG	7350
AATCCAATGT	GTCTTCTGCC	TTGGCGGAGC	TGCCCCTAA	GACCTTCGGT	7400
AGCTCCGGAT	CGTCGGCCGT	TGATAGCGGC	ACGGCGACCG	CCCTTCCCTGA	7450
CCTGGCCTCC	GACGACGGTG	ACAAAGGATC	CGACGTTGAG	TGCTACTCCT	7500
CCATGCCCCC	CCTTGAAGGG	GAGCCGGGGG	ACCCCGATCT	CAGCGACGGG	7550
TCTTGGTCTA	CCGTGAGTGA	GGAGGCTAGT	GAGGATGTGG	TCTGCTGCTC	7600

FIG. 14D

HC-J4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AATGTCTTAT	ACGTGGACAG	GCGCCCTGAT	CACGCCATGC	GCTGCGGAGG	7650
AAAGTAAAGCT	GCCCATCAAC	CCGTTGAGCA	ACTCTTTTGCT	GCGTCACCAC	7700
AACATGGTCT	ACGCCACAAC	ATCCCGCAGC	GCAAGCCTOC	GGCAGAAGAA	7750
GGTCACCTTT	GACAGATTGC	AAGTCTTGGA	TGATCATTAC	CGGGACGTAC	7800
TCAAGGAGAT	GAAGGCGAAG	GCGTCCACAG	TTAAGGCTAA	GCTTCTATCT	7850
ATAGAGGAGG	CCTGCAAGCT	GACGCCCCCA	CATTGGGCA	AATCCAAATT	7900
TGGCTATGGG	GCAAAGGACG	TCCGGAACT	ATCCAGCAGG	GCCGTTAACC	7950
ACATCCGCTC	CGTGTGGGAG	GACTTGCTGG	AAGACACTGA	AACACCAATT	8000
GACACCACCA	TCATGGCAAA	AAGTGGGTT	TTCTGGGTCC	AACCAGAGAA	8050
GGGAGGCCGC	AAGCCAGCTC	GCCTTATCGT	ATTCCCAGAC	CTGGGAGTTC	8100
GTGTATGCGA	GAAGATGGCC	CTTTACGACG	TGGTCTCCAC	CCTTCTCAG	8150
GCCGTGATGG	GCTCCTCATA	CGGATTTCAA	TACTCCCCCA	AGCAGCGGGT	8200
CGAGTTCTTG	GTGAATACCT	GGAAATCAAA	GAAATGCCCT	ATGGGCTTCT	8250
CATATGACAC	CCGCTGTTTT	GACTCAACGG	TCACTGAGAG	TGACATTGGT	8300
GTTGAGGAGT	CAATTTACCA	ATGTTGTGAC	TTGGCCCCCG	AGGCCAGACA	8350
GGCCATAAGG	TCGCTCACAG	AGCGGCTTTA	CATCGGGGGT	CCCTGACTA	8400
ACTCAAAGG	GCAGAACTGC	GGTTATCGCC	GGTGCCGGGC	AAGTGGCGTG	8450
CTGACGACTA	GCTGCGGTAA	TACCTCACA	TGTTACTTGA	AGGCCACTGC	8500
AGCCTGTGCA	GCTGCAAAGC	TCCAGGACTG	CACGATGCTC	GTGAACGGAG	8550
ACGACCTTGT	CGTTATCTGT	GAAAGCGCGG	GAACCCAGGA	GGATGCGGGC	8600
GCCCTACGAG	CCTTCACGGA	GGCTATGACT	AGGTATTTCG	CCCCCCCCCG	8650
GGATCCGGCC	CAACCAGAAT	ACGACCTGGA	GCTGATAACA	TCATGTTCTT	8700
CCAATGTGTC	AGTCGCGCAC	GATGCATCTG	GCAAAAGGGT	ATACTACCTC	8750
ACCCGTGACC	CCACCACCCC	CCTTGACAGG	GCTGCGTGGG	AGACAGCTAG	8800
ACACACTCCA	ATCAACTCTT	GGCTAGGCAA	TATCATCATG	TATGCGCCCA	8850
CCCTATGGGC	AAGGATGATT	CTGATGACTC	ACTTTTCTCT	CATCCTTCTA	8900
GCTCAAGAGC	AACCTGAAAA	AGCCCTGGAT	TGTCAGATCT	ACGGGGCTTG	8950
CTACTCCATT	GAGCCACTTG	ACCTACCTCA	GATCATTGAA	CGACTCCATG	9000
GTCTTAGCGC	ATTTACACTC	CACAGTTACT	CTCCAGGIGA	GATCAATAGG	9050
GTGGCTTCAT	GCCTCAGGAA	ACTTGGGGTA	CCACCTTGC	GAACCTGGAG	9100
ACATCGGGCC	AGAAGTGTCC	GCGCTAAGCT	ACTGTCCAG	GGGGGAGGG	9150
CCGCCACTTG	TGGCAGATAC	CTCTTTAACT	GGGCAGTAAG	GACCAAGCTT	9200
AAACTCACTC	CAATCCCGGC	CGCGTCCAG	CTGGACTTGT	CTGGCTGGTT	9250
CGTCGCTGGT	TACAGCGGGG	GAGACATATA	TCACAGCCTG	TCTGTGCCC	9300
GACCCCGCTG	GTTCGCTTG	TGCCACTCC	TACTTTCTGT	AGGGGTAGGC	9350
ATTTACCTGC	TCCCCAACCG	ATGAACGGGG	AGCTAACCCAC	TCCAGGCCCT	9400
AAGCCATTTC	CIGTTTTTTT	TTTTTTTTTT	TTTTTTTTTT	TCTTTTTTTT	9450
TTTCTTTCTT	TTCCTTCTTT	TTTTCTTTTC	TTTTTCCCTT	CTTTAATGGT	9500

FIG. 14E

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GGCTCCATCT	TAGCCCTAGT	CACGGCTAGC	TGTGAAAGGT	CCGTGAGCCG	9550
CATGACTGCA	GAGAGTGCTG	ATACTGGCCT	CTCTGCAGAT	CATGT	9595

FIG. 14F

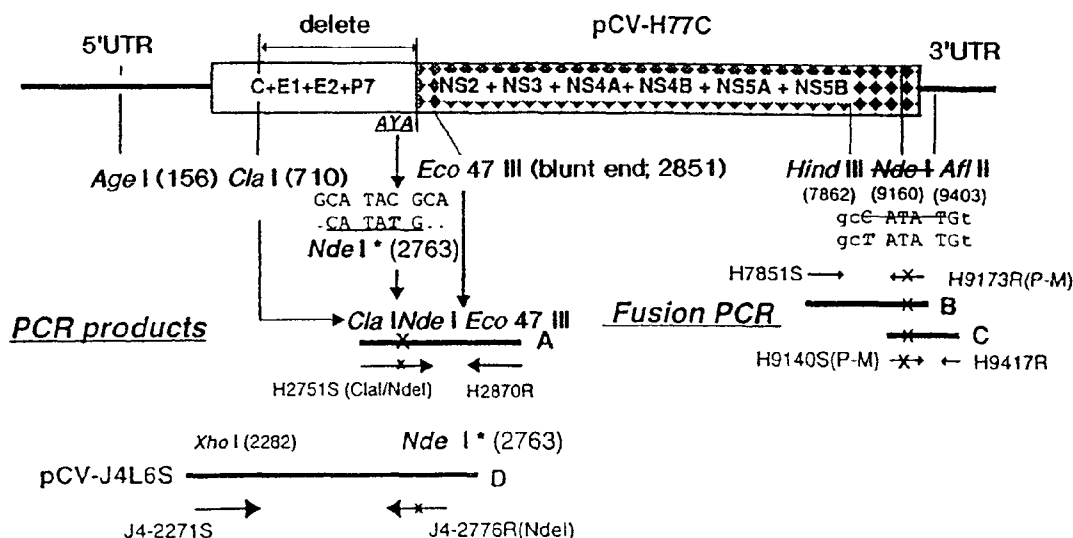
10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
MSINPKPQRK	TKRNINRRPQ	DVKFPGGGQI	VGGVYLLPRR	GPRLGVRATR	50
KASERSQPRG	RRQPIPKARR	PEGRAWAQPG	YPWPLYGNEG	LQWAGWLLSP	100
RGSRPSWGPT	DPRRRSRNLG	KVIDTLITCGF	ADLMGYIPLV	GAPLGGAARA	150
LAHGVRLVED	GVNYATGNLP	GCSFSIFLLA	LLSCLTIPAS	AYEVRNVSGI	200
YHVINDCSNS	SIVYEAADVI	MHTPGCVPCV	QEGNSSRCW	ALTPILAARN	250
ASVPTTITRR	HVDLLVGTA	FCSAMYVGD	CGSIFLVSQ	FTFSPRRHET	300
VQDCNCSTYP	GHVSGHRMAW	DMMNWSPTT	ALVVSQLLRI	PQAVVDMVAG	350
AHWGVLAGLA	YYSMVGNWAK	VLIVALLFAG	VDGETHITGR	VAGHITSGFT	400
SLFSSGASQK	IQLVNTINGSW	HINRTALN	DSLQITGFFAA	LFYAHKFNSS	450
GCPERMASCR	PIDWFAQGWG	PITYTKFNSS	DQRPYCWHYA	PRPGWVPAS	500
QVCGPVYCF	PSPVVVGITD	RSGVPTYSWG	ENETDMLLN	NIRPPQGNWF	550
GCTWMNSTGF	TKTCGGPPCN	IGGVGNRILI	CPTDCFRKHP	EATYTKCGSG	600
PWLTPRCLVD	YPYRLWHYPC	TLNFSIFKVR	MYVGGVEHRL	NAACNWIRGE	650
RCNLEDRDRS	ELSPLLLSTT	EWQILECAFT	TLPALSTGLI	HLHQNTVDVQ	700
YLYGVGSAFV	SFAIKWEYIL	LLFLLILADAR	VCACLWMMLL	IAQAEAALEN	750
LVLVNAASVA	GAHGILSFLV	FFCAAWYIKG	RLAPGAAYAF	YGVWPLLLLL	800
LALPPRAYAL	DREMAASCGG	AVLVGLVFLT	LSPYYKVFLT	RLIWWLQYFI	850
TRAEAHMQW	VPPLNVRGGR	DAIILLITCAV	HPELIFDITK	LLLAILGPLM	900
VLQAGITRVP	YFVRAQGLIR	ACMLVRKVAG	GHYVQMVFMK	LGALTGTIVY	950
NHLTPLRDWA	HAGLRDLAVA	VEPVVFSAME	TKVTTWGADT	AACGDIILGL	1000
PVSARRGKEI	FLGPADSLG	QGWRLAPIT	AYSQQTRGVL	GCIITSLTGR	1050
DKNQVEGEVQ	VVSTATQSFL	ATCINGVCWT	VYHGAGSKTL	AGPKGPITQM	1100
YTNVDLDLVG	WQAPPGARSM	TPCSCGSSDL	YLVTRHADVI	FVRRRGDSRG	1150
SLLSRPVSY	LKGSSGGPLL	CPSGHVGVF	RAAVCTRGA	KAVDFIPVES	1200
METIMRSPVF	TDNSTPPAVP	QTFQVAHLHA	PTGSGKSTKV	PAAYAAQGYK	1250
VLVLNPSVAA	TLGFGAYMSK	AHGIDPNIRT	GVRTTTTIGGS	ITYSTYGKFL	1300
ADGGCSCGAY	DIIICDECHS	TDSTTILGIG	TVLDQAETAG	ARLVLATAT	1350
PPGSVTVPHP	NIEEIGLSNN	GEIPFYGKAI	PIEAIKGRH	LIFCHSKKKC	1400
DELAAKLTGL	GLNAVAYYRG	LDVSVIPIG	DVVVATDAL	MITGFTGDFDS	1450
VIDCNICVTQ	TVDFSLDPTF	TIETTTVPQD	AVSRSQRRGR	TGRGRSGIYR	1500
FVTPGERPSG	MFDSSVLCEC	YDAGCAWYEL	TPAETSVRLR	AYLNTPLPLV	1550
QQDHLEFWES	VFTGLTHIDA	HFLSQTKQAG	DNFPYLVAYQ	ATVCARAQAP	1600
PPSWDQMKC	LIRLKPTLHG	PTPLLYRIGA	VQNEVILTHP	ITKYIMACMS	1650
ADLEVVTSIW	VLVGGVLAAL	AAYCLITGVS	VTVGRIILSG	KPAVVPDREV	1700
LYQEFDEMEE	CASQLFYIEQ	GMQLAEQFKQ	KALGLLQTAT	KQAEAAAPVV	1750
ESKWRALET	WAKHMANFIS	GIQYLAGLST	LPGNPAIASL	MAFTASITSP	1800
LITQNTLLFN	ILGGWAAQL	APPSAASAFV	GAGIAGAAVG	SIGLGKVLVD	1850
ILAGYGAGVA	GALVAFKVM	GEVPSTEDLV	NLLPAILSPG	ALVGVVCAA	1900

FIG. 14G

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ILRRHVGEGE	GAVQWMNRLI	AFASRGNHVS	PIHYVPESDA	AARVIQILSS	1950
LITTIQLLKRL	HQWINEDCST	PCSGSWLRDV	WDWICTVLTD	FKIWLQSKLL	2000
PRLPGVPFLS	CQRGYKGVWR	GDGIMQTTCP	CGAQLAGHVK	NGSMRIVGPR	2050
TCSNIWHGIF	PINAYTTGPC	TPSPAIFYSR	ALWRVAAEEY	VEVIRVGDFH	2100
YVTGMTIDNV	KCPQVPAPE	FFTEVDGVRL	HRYPACKPL	LREDVTFQVG	2150
LNQYLVGSOL	PCEPEPDVIV	LTSMLTDPST	TTAETAKRRL	ARGSPPSLAS	2200
SSASQLSAPS	LKATCTIHHD	SPDADLLEAN	LLWRQEMGGN	ITRVESENKV	2250
VILDSFEPLH	AEGDEREISV	AAETLRKSRK	FPSALPIWAR	PDYNPPLLES	2300
WKDPDYVPFV	VHGCPLPTIK	APPIPPPRRK	RIVVLITESV	SSALAEIATK	2350
TFGSSGSSAV	DSGTATALPD	LASDDGDKGS	DVESYSSMPP	LEGEFGDPLD	2400
SDGSWSIVSE	EASEDVVCCS	MSYTWIGALI	TPCAAEEESKL	PINPLSNSLL	2450
RHHNMVYATT	SRSASLRQKK	VTFDRLQVLD	DHYRDVLKEM	KAKASTVKAK	2500
LLSIEEACKL	TPPHSAKSKF	GYGAKIVRNL	SSRAVNHIRS	VWEDLLEDTE	2550
TPIDTTIMAK	SEVFCVQPEK	GGRKPARLIV	FPDLGVRVCE	KMALYDVVST	2600
LPQAVMGSSY	GFQYSPKQRV	EFLVNIWVSK	KCPMGFSYDT	RCFDSTVIES	2650
DIRVEESTYQ	OCDLAPEARQ	AIRSLTERLY	IGGPLTNSKG	QNOGYRRORA	2700
SGVLTTSCGN	TLTCYLKATA	ACRAAKLQDC	TMLVNGDDL	VICESAGTQE	2750
DAAALRAFTE	AMIRYSAPFG	DPPQPEYDLE	LITSCSSNVS	VAHDASGKRV	2800
YYLTRDPTTP	LARAAWETAR	HTPINSWLGN	IIMYAPTLWA	RMILMIHFFS	2850
ILLAQEQLK	ALDCQIYGAC	YSIEPLDLPQ	IIEHLHGLSA	FTLHSYSPGE	2900
INRVASCLRK	LGVPLRTWR	HRARSVRACL	LSQGGRAATC	GRYLEFNWAVR	2950
TKLKLTPIPA	ASQLDLGWFF	VAGYSGGDIY	HSLSRARPRW	FPLCLLLLSV	3000
GVGIYLLPNR					3010

FIG. 14H

#2. Strategy for constructing chimeric clone of HCV (pH77CV-J4) which contains the nonstructural region of strain H77 and the structural region of strain HC-J4



1. Fragment A, B, C and D ; PCR amplification from pCV-H77C or pCV-J4L6S
 - Fragment A ; additional *Cla* I site, artificial *Nde* I site induced by a single mutation (C→T at nt 2765 of H77C) and authentic *Eco*47 III site
 - Fragment B and C ; eliminated *Nde* I site by a single mutation within the primers (C→T at nt 9158 of H77C) , and fusion PCR with both fragments
 - Fragment D ; artificial *Nde* I site induced by 2 point mutations within the primer (T→A at nt 2762 and C→T at nt 2765 of J4L6S)
2. TA cloning of PCR products
3. Sequence analysis
4. Cloning of Fragment A (*Cla* I-*Eco* 47III) and Fragment B/C (*Hind* III-*Afl* II) with correct sequence into pCV-H77C
5. Complete sequence analysis of new cassette vector [pH77CV], into which the structural regions of different genotypes can be inserted.
6. Cloning of Fragment-Age I/*Xho* I (cut out from pCV-J4L6S) and Fragment D (*Xho* I-*Nde* I) with correct sequence into the new cassette vector ; 3 piece ligation
7. Complete sequence analysis of 1a+1b chimera [pH77CV-J4]
8. *In vitro* transcription (within 24 hours of inoculation)
9. Percutaneous intrahepatic transfection into chimpanzee

FIG. 15

GCCAGCCCCC	TGATGGGGGC	GACACTCCAC	CATGAATCAC	TCCCTGTGTA	50
GGAACTACTG	TCTTCACGCA	GAAAGCGTCT	AGCCATGGCG	TTAGTATGAG	100
TGTGTGTCAG	CCTCCAGGAC	CCCCCTCCC	GGGAGAGCCA	TAGTGGTCTG	150
CGGAACCGGT	GAGTACACCG	GAATTGCCAG	GACGACCGGG	TCCTTTCTTG	200
GATCAACCCG	CTCAATGCCT	GGAGATTTGG	GCGTGCCCCC	GCGAGACTGC	250
TAGCCGAGTA	GTGTTGGGTC	GCGAAAGGCC	TTGTGGTACT	GCCTGATAGG	300
GTGCTTGCGA	GTGCCCCGGG	AGGTCTCGTA	GACCGTGCAC	CATGAGCAAG	350
AATCCTAAAC	CTCAAAGAAA	AACCAAAGCT	AACACCAACC	GGGCCCCACA	400
GGAGTCAAG	TTCCCCGGCG	GTGGTCAGAT	CGTTGGTGGG	GTTTACCTGT	450
TGCCCCGCGG	GGGCCCCAGG	TTGGGTGTGC	GCGCGACTAG	GAAGGCTTCC	500
GAGCGGTCCG	AACCTCGTGG	AAGGCGACAA	CCTATCCCAA	AGGCTCGCCG	550
ACCCGAGGGC	AGGGCCTGGG	CTCAGCCCCG	GTACCTTTGG	CCCCCTCATG	600
GCAATGAGGG	CCTGGGGTGG	GCAGGATGGC	TCCTGTCAAC	CCGCGGCTCC	650
CGGCCTAGTT	GGGGCCCCAC	GGACCCCCCG	CGTAGGTCCG	GTAACTTGGG	700
TAAGGTCATC	GATACCCCTA	CATGCGGCTT	CGCCGATCTC	ATGGGGTACA	750
TTCCGCTCGT	CGGCGCCCCC	CTAGGGGGCG	CTGCCAGGGC	CTTGGCACAC	800
GGTGTCCGGG	TTCTGGAGGA	CGGCGTGAAC	TATGCAACAG	GGAACTTGCC	850
CGGTGTCTCT	TTCTCTATCT	TCCTCTTGGC	TCTGCTGTCC	TGTTTGACCA	900
TCCCAGCTTC	CGCTTATGAA	GTGCCCAACG	TGTCCGGGAT	ATACCATGTC	950
ACGAACGACT	GCTCCAACCT	AAGCATTGTG	TATGAGGCAG	CGGACGTGAT	1000
CATGCATACT	CCCGGGTGGG	TGCCCTGTGT	TCAGGAGGGT	AACAGCTCCC	1050
GTGTCTGGGT	AGCGCTCACT	CCCACGCTCG	CGGCCAGGAA	TGCCAGCGTC	1100
CCCACTACGA	CAATACGACG	CCACGTCCAC	TTGCTCGTTG	GGACGGCTGC	1150
TTTCTGCTCC	GCTATGTACG	TGGGGGATCT	CTGCCGATCT	ATTTTCTCTG	1200
TCTCCAGCT	GTTCAACCTC	TGCCCTCGCC	GGCATGAGAC	AGTGCAGGAC	1250
TGCAACTGCT	CAATCTATCC	CGGCCATGTA	TCAGGTCAAC	GCATGGCTTG	1300
GGATATGATG	ATGAACTGGT	CACCTACAAC	AGCCCTAGTG	GTGTCCGAGT	1350
TGCTCCGGAT	CCCACAAGCT	GTGTGGGACA	TGGTGGCCGG	GGCCCACTGG	1400
GGAGTCTTGG	CGGGCCTTGC	CTACTATTCC	ATGGTAGGGA	ACTGGGCTAA	1450
GGTTCTGATT	GTGGCGCTAC	TCTTTGCCGG	CGTTGACGGG	GAGACCCACA	1500
CGACGGGGAG	GGTGGCCGGC	CACACCACCT	CCGGGTTCAC	GTCCCTTTTC	1550
TCATCTGGGG	CGTCTCAGAA	AATCCAGCTT	GTGAATACCA	ACGGCAGCTG	1600
GCACATCAAC	AGGACTGCCC	TAAATTGCAA	TGACTCCCTC	CAAACCTGGT	1650
TCTTTGCCCG	GCTGTTTAC	GCACACAAGT	TCAACTCGTC	CGGGTGCCCG	1700
GAGCGCATGG	CCAGCTGCCG	CCCCATTGAC	TGGTTCCGCC	AGGGGTGGGG	1750
CCCCATCACC	TATACTAAGC	CTAACAGCTC	GGATCAGAGG	CCTTATTGCT	1800

FIG. 16A

GGCATTACGC	GCCTCGACCG	TGIGGTGTG	TACCCGCGTC	GCAGGTGTGT	1850
GGTCCAGTGT	ATTGTTTCAC	CCCAAGCCCT	GTGTGTGGTG	GGACCAACGA	1900
TGCTTCCGGT	GTCCCTACGT	ATAGCTGGGG	GGAGAATGAG	ACAGACGTGA	1950
TGCTCCTCAA	CAACACGGT	CCGCCACAAG	GCAACTGGTT	CGGCTGTACA	2000
TGGATGAATA	GTACTGGGTT	CACTAAGACG	TGCGGAGGTC	CCCCGTGTAA	2050
CATCGGGGGG	GTGGTTAACC	GCACCTTGAT	CTGCCCCACG	GA CTGCTTCC	2100
GGAAGCACCC	CGAGGCTACT	TACACAAAAT	GTGGCTCGGG	GCCCTGGTTG	2150
ACACCTAGGT	GCCTAGTAGA	CTACCCATAC	AGGCTTTGGC	ACTACCCCTG	2200
CACICTCAAT	TTTTCCATCT	TTAAGGTTAG	GATGTATGTG	GGGGGCGTGG	2250
AGCACAGGCT	CAATGCCGCA	TGCAATTGGA	CTCGAGGAGA	GCGCTGTAAAC	2300
TTGGAGGACA	GGGATAGGTC	AGAACTCAGC	CCGCTGCTGC	TGTCTACAAC	2350
AGAGTGGCAG	ATACTGCCCT	GTGCTTTTCAC	CACCCCTACCG	GCTTTATCCA	2400
CTGGTTTGAT	CCATCTCCAT	CAGAACATCG	TGGACGTGCA	ATACTGTAC	2450
GGTGTAGGGT	CAGCGTTTGT	CTCCTTTTGA	ATCAAATGGG	AGTACATCCT	2500
GTGTCTTTTC	CTTCTCCTGG	CAGACGGCGG	CGTGTGTGCC	TGCTTGTGGA	2550
TGATGCTGCT	GATAGCCCAG	GCTGAGGCCG	CCTTAGAGAA	CTTGGTGGTC	2600
CTCAATGCCG	CGTCCGTGCC	CGGAGCGCAT	GGTATTTCTCT	CCTTTCTTGT	2650
GTTCCTTCTG	GCCGCCCTGGT	ACATTAAAGG	CAGGCTGGCT	CCTGGGGCGG	2700
CGTATGCTTT	TTATGGCGTA	TGGCCGCTGC	TCTGCTCCT	ACTGGCGTFA	2750
CCACCACGAG	CATATGCACT	GGACACGGAG	GTGGCCCGGT	CGTGTGGCGG	2800
CGTGTCTCTT	GTCCGGTTAA	TGGCGCTGAC	TCTGTGCGCA	TATTACAAGC	2850
GCTATATCAG	CTGGTGCA TG	TGGTGGCTTC	AGTATTTTCT	GACCAGAGTA	2900
GAAGCGCAAC	TGCACGTGTG	GGTTCCCCCC	CTCAACGTCC	GGGGGGGGCG	2950
CGATGCCGTC	ATCTTACTCA	TGTGTGTAGT	ACACCCGACC	CTGGTATTTG	3000
ACATCACCAA	ACTACTCCTG	GCCATCTTGG	GACCCCTTTG	GATTCTTCAA	3050
GCCAGTTTGC	TTAAAGTCCC	CTACTTCGTG	CGCGTTCAAG	GCCTTCTCCG	3100
GATCTGCCCG	CTAGCCCGGA	AGATAGCCCG	AGGTCATTAC	GTGCAAAATG	3150
CCATCATCAA	GTTAGGGGGG	CTTACTGGCA	CCTATGTGTA	TAAACATCTC	3200
ACCCCTCTTC	GAGACTGGGC	GCACAACGGC	CTGCGAGATC	TGGCCGTGGC	3250
TGTGGAACCA	GTCGTCTTCT	CCCGAATGGA	GACCAAGCTC	ATCACGTGGG	3300
GGGCAGATAC	CGCCCGGTGC	GGTGACATCA	TCAACGGCTT	GCCCGTCTCT	3350
GCCCGTAGGG	GCCAGGAGAT	ACTGCTTGGG	CCAGCCGACG	GAATGGTCTC	3400
CAAGGGGTGG	AGGTTGCTGG	CGCCCATCAC	GCGGTACGCC	CAGCAGACGA	3450
GAGGCCTCCT	AGGGTGTATA	ATCACCAGCC	TGACTGGCCG	GGACAAAAAC	3500
CAAGTGAGAG	GTGAGGTCCA	GATCGTGTCA	ACTGCTAACC	AAACCTTCCT	3550
GGCAACGTGC	ATCAATGGGG	TATGCTGGAC	TGTCTACCAC	GGGGCCGGAA	3600

FIG. 16B

CGAGGACCAT	CGCATCACCC	AAGGGTCCTG	TCATCCAGAT	GTATACCAAT	3650
GTGGACCAAG	ACCTTGTTGGG	CTGGCCCCGT	CCTCAAGGTT	CCCCCTCATT	3700
GACACCCTGT	ACCTGCGGCT	CCTGGAGCCT	TTACCTGGTC	ACGAGGCACG	3750
CCGATGTCAT	TCCCCGTGCGC	CGGCGAGGTG	ATAGCAGGGG	TAGCCTGCTT	3800
TCGCCCCGGC	CCATTTTCTA	CTTGAAAGGC	TCCTCGGGGG	GTCCGCTGTT	3850
GTGCCCCGGG	GGACACGCGG	TGGGCTTATT	CAGGGCCGGG	GTGTGCACCC	3900
GTGGAGTGGC	TAAAGCGGTG	GACTTTTATC	CTGTGGAGAA	CCTAGGGACA	3950
ACCATGAGAT	CCCCGGTGT	CACGGACAAC	TCCTCTCCAC	CAGCAGTGCC	4000
CCAGAGCTTC	CAGGTGGCCC	ACCTGCATGC	TCCCACCGGC	AGCGGTAAAG	4050
GCACCAAGGT	CCCCGCTGGG	TACGCAGCCC	AGGGCTACAA	GGTGTGTTGG	4100
CTCAACCCCT	CTGTGTGCTG	AACGCTGGGC	TTTGGTGTCT	ACATGTCCAA	4150
GGCCCATGGG	GTGTATCCTA	ATATCAGGAC	CGGGTGAGAA	ACAATTACCA	4200
CTGGCAGCCC	CATCACGTAC	TCCACCTACG	GCAAGTTTCT	TGCCGACGGC	4250
GGGTGCTCAG	GAGGTGCTTA	TGACATAATA	ATTTGTGACG	AGTGCCACTC	4300
CACGGATGCC	ACATCCATCT	TGGGCATCGG	CACTGTCTCT	GACCAAGCAG	4350
AGACTGCGGG	GGCGAGACTG	GTTGTGCTCG	CCACTGCTAC	CCCTCCGGGC	4400
TCGGTCACTG	TGTCCCATCC	TAACATCGAG	GAGGTTGCTC	TGTCCACCAC	4450
CGGAGAGATC	CCCTTTTACG	GCAAGGCTAT	CCCCCTCGAG	GTGATCAAGG	4500
GGGGAAGACA	TCTCATCTTC	TGCCACTCAA	AGAAGAAGTG	CGACGAGCTC	4550
GCCGCGAAGC	TGGTCCGATT	GGCATCAAT	GCCGTGGCCT	ACTACCGCGG	4600
TCTTGACGTG	TCTGTCTATC	CGACCAGCGG	CGATGTTGTC	GTCGTGTCCA	4650
CCGATGCTCT	CATGACTGGC	TTTACCGGGG	ACTTCGACTC	TGTGATAGAC	4700
TGCAACACGT	GTGTCACTCA	GACAGTCGAT	TTCAGCCTTG	ACCCTACCTT	4750
TACCATTGAG	ACAACCACGC	TCCCCCAGGA	TGCTGTCTCC	AGGACTCAAC	4800
GCCGGGGCAG	GACTGGCAGG	GGGAAGCCAG	GCATCTATAG	ATTTGTGGCA	4850
CCGGGGGAGC	GCCCCCTCCG	CATGTTTCGAC	TGTTCCGTCC	TCTGTGAGTG	4900
CTATGACCGG	GGCTGTCTCT	GGTATGAGCT	CACGCCCCGC	GAGACTACAG	4950
TTAGGCTACG	AGCGTACATG	AACACCCCGG	GGCTTCCCGT	GTGCCAGGAC	5000
CATCTTGAAT	TTTGGGAGGG	CGTCTTTTAC	GGCTCACTC	ATATAGATGC	5050
CCACTTTTTTA	TCCCAGACAA	AGCAGAGTGG	GGAGAACTTT	CCTTACCTGG	5100
TAGCGTACCA	AGCCACCGTG	TGCGCTAGGG	CTCAAGCCCC	TCCCCCATCG	5150
TGGGACCAGA	TGTGGAAGTG	TTTGATCCGC	CTTAAACCCA	CCCTCCATGG	5200
GCCAACACCC	CTGCTATACA	GACTGGGGGC	TGTTTCAGAAT	GAAGTCACCC	5250
TGACGCACCC	AATCACCAAA	TACATCATGA	CATGCATGTC	GGCCGACCTG	5300
GAGGTGCTCA	CGAGCACCTG	GGTGCTCGTT	GGCGGGTCC	TGGCTGCTCT	5350
GGCCGCGTAT	TGCCGTGTCAA	CAGGCTCGGT	GGTCATAGTG	GGCAGGATCG	5400

FIG. 16C

TCTTGTCGGG	GAAGCCGGCA	ATTATACCTG	ACAGGGAGGT	TCTCTACCAG	5450
GAGTTGATG	AGATGGAAGA	GTGCTCTCAG	CACTTACOGT	ACATCGAGCA	5500
AGGGATGATG	CTCGCTGAGC	AGTTCAAGCA	GAAGGCOCTC	GGCCTCCTGC	5550
AGACCGCGTC	CCGCCATGCA	GAGGTATATCA	CCCCTGCTGT	CCAGACCAAC	5600
TGGCAGAAAC	TCGAGGTCTT	TTGGGCGAAG	CACATGTGGA	ATTTTCATCAG	5650
TGGGATACAA	TACTTGGCGG	GCCTGTCAAC	GCTGCCCTGGT	AACCCCGCCA	5700
TTGCTTCATT	GATGGCTTTT	ACAGCTGCCG	TCACCAGCCC	ACTAACCACT	5750
GGCCAAACCC	TCCTCTTCAA	CATATTGGGG	GGGTGGGTGG	CTGCCCAGCT	5800
CGCCGCCCCC	GGTCCCGCTA	CTGCCCTTTGT	GGGTGCTGGC	CTAGCTGGCG	5850
CCGCCATCGG	CAGCGTTGGA	CTGGGGAAGG	TCCTCGTGGA	CATTCTTGCA	5900
GGGTATGGCG	CGGGCGTGGC	GGGAGCTCTT	GTAGCAITCA	AGATCATGAG	5950
CGGTGAGGTC	CCCTCCACGG	AGGACCTGGT	CAATCTGCTG	CCCGCCATCC	6000
TCTCGCCTGG	AGCCCTTGTA	GTCCGTGTGG	TCIGCGCAGC	AATACTGCGC	6050
CGGCACGTTG	GCCCCGGCGA	GGGGGCAGTG	CAATGGATGA	ACCGGCTAAT	6100
AGCCTTCGCC	TCCCCGGGGG	ACCATGTTTC	CCCCACGCAC	TACGTGCCCG	6150
AGAGCGATGC	AGCCGCCCGC	GTCACTGCCA	TACTCAGCAG	CCTCACTGTA	6200
ACCCAGCTCC	TGAGGCGACT	GCATCAGTGG	ATAAGCTCCG	AGTGTACCAC	6250
TCCATGCTCC	GGTTCTTGGC	TAAGGGACAT	CTGGGACTGG	ATATGCGAGG	6300
TGCTGAGCGA	CTTTAAGACC	TGGCTGAAAG	CCAAGCTCAT	GCCACAACCTG	6350
CCTGGGATTG	CCTTTGTGTC	CTGCCAGCGC	GGGTATAGGG	GGGTCTGGCG	6400
AGGAGACGGC	ATTATGCACA	CTCGCTGCCA	CTGTGGAGCT	GAGATCACTG	6450
GACATGTCAA	AAACGGGACG	ATGAGGATCG	TOGGTCCTAG	GACCTGCAGG	6500
AACATGTGGA	GTGGGACGTT	CCCCATTAAAC	GCCTACACCA	CGGGCCCCCTG	6550
TACTCCCCCTT	CCTGCCGCCA	ACTATAAGTT	CGCGCTGTGG	AGGGTGTCTG	6600
CAGAGGAATA	CGTGGAGATA	AGGCGGGTGG	GGGACTTCCA	CTACGTATCG	6650
GGTATGACTA	CTGACAATCT	TAAATGCCCG	TGCCAGATCC	CATCGCCCCGA	6700
ATTTTTTCACA	GAATTGGACG	GGGTGCGCCT	ACACAGGTTT	GCGCCCCCTT	6750
GCAAGCCCTT	GCTGCGGGAG	GAGGTATCAT	TCAGAGTAGG	ACTCCACGAG	6800
TACCCGGTGG	GGTCCGAATT	AOCCTGCGAG	CCCGAACCGG	ACGTAGCCGT	6850
GTTGACGTCC	ATGCTCACTG	ATCCCTCCCA	TATAACAGCA	GAGGCGGCCG	6900
GGAGAAGGTT	GGCGAGAGGG	TCACCCCTTT	CTATGGCCAG	CTCCTCGGCT	6950
AGCCAGCTGT	CCGCTCCATC	TCTCAAGCCA	ACTTGCACCG	CCAACCATGA	7000
CTCCCCCTGAC	GCCGAGCTCA	TAGAGGCTAA	CCTCCTGTGG	AGGCAGGAGA	7050
TGGGCGGCAA	CATCACCAGG	GTTGAGTCAG	AGAACAAAGT	GGTGATTCTG	7100
GACTCCTTCG	ATCCGCTTGT	GGCAGAGGAG	GATGAGCGGG	AGGTCTCCGT	7150
ACCTGCAGAA	ATTCTGCGGA	AGTCTCGGAG	ATTGCCCCGG	GCCCTGCCCC	7200

FIG. 16D

TCTGGGCGCG	GCCGGACTAC	AACCCCCCGC	TAGTAGAGAC	GTGGAAAAG	7250
CCTGACTACG	AACCACCTGT	GGTCCATGGC	TGCCCCGTAC	CACCTCCACG	7300
GTCCCCCTCT	GTGCTCCGC	CTCGGAAAAA	GCGTACGGTG	GTCCTCACCG	7350
AATCAACCTT	ATCTACTGCC	TTGGCCGAGC	TTGCCACCAA	AAGTTTTGGC	7400
AGCTCCTCAA	CTTCCGGCAT	TACGGGGGAC	AATACGACAA	CATCCTCTGA	7450
GCCCCCCCC	TCTGGCTGCC	CCCCCGACTC	CGACGTTGAG	TCCTATTCTT	7500
CCATGCCCCC	CCTGGAGGGG	GAGCCTGGGG	ATCCGGATCT	CAGCGACGGG	7550
TCATGGTCCA	CGGTCAGTAG	TGGGGCCGAC	ACGGAAGATG	TCGTGTGCTG	7600
CTCAATGTCT	TATTCTTGA	CAGGCGCACT	CGTCACCCCG	TGCGCTGCGG	7650
AAGAACA AAA	ACTGCCCATC	AACGCACTGA	GCAACTCGTT	GCTACGCCAT	7700
CACAATCTGG	TGTATTCCAC	CACTTCAAGC	AGTGCTTGCC	AAAGGCAGAA	7750
GAAAGTCACA	TTTGACAGAC	TGCAAGTTCT	GGACAGCCAT	TACCAGGACG	7800
TGCTCAAGGA	GGTCAAAGCA	GCGGCGTCAA	AAGTGAAGGC	TAACTTGCTA	7850
TCCGTAGAGG	AAGCTTGCAG	CCTGACGCCC	CCACATTGAG	CCAAATCCAA	7900
GTTTGGCTAT	GGGGCAAAAG	ACGTCCGTTG	CCATGCCAGA	AAGGCCGTAG	7950
CCCACATCAA	CTCCGTGTGG	AAAGACCTTC	TGGAAGACAG	TGTAAACACCA	8000
ATAGACACTA	CCATCATGGC	CAAGAACGAG	GTTTTCTGCG	TTCAGCCTGA	8050
GAAGGGGGGT	CGTAAGCCAG	CTCGTCTCAT	CGTGTTCCCC	GACCTGGGCG	8100
TGCGCGTGTG	CGAGAAGATG	GCCCTGTACG	ACGTGGTTAG	CAAGCTCCCC	8150
CTGGCCGTGA	TGGGAAGCTC	CTACCGATTG	CAATACTCAC	CAGGACAGCG	8200
GGTTGAATTG	CTCGTGCAAG	CGTGGGAAGT	CAAGAAGACC	CCGATGGGGT	8250
TCTCGTATGA	TACCCGCTGT	TTTGACTCCA	CAGTCACTGA	GAGCGACATC	8300
CGTACGGAGG	AGGCAATTTA	CCAATGTTGT	GACCTGGACC	CCCAAGCCCG	8350
CGTGGCCATC	AAGTCCCTCA	CTGAGAGGCT	TTATGTTGGG	GGCCCTCTTA	8400
CCAATTCAAG	GGGGGAAAC	TGCGGCTACC	GCAGGTGCCG	CGCGAGCCGC	8450
GTACTGACAA	CTAGCTGTGG	TAACACCCCT	ACTTGCTACA	TCAAGGCCCG	8500
GGCAGCCTGT	CGAGCCGCAG	GGCTCCAGGA	CTGCACCATG	CTCGTGTGTG	8550
GCGACGACTT	AGTCGTTATC	TGTGAAAGTG	CGGGGGTCCA	GGAGGACGGG	8600
GCGAGCCTGA	GAGCCTTCAC	GGAGGCTATG	ACCAGGTAAT	CCGCCCCCCC	8650
CGGGGACCCC	CCACAACCAG	AATACGACTT	GGAGCTTATA	ACATCATGCT	8700
CCTCCAACGT	GTCAGTCGCC	CACGACGGCG	CTGGAAAGAG	GGTCTACTAC	8750
CTTACCCGTG	ACCTTACAAC	CCCCCTCGCG	AGAGCCGCGT	GGGAGACAGC	8800
AAGACACACT	CCAGTCAATT	CCTGGCTAGG	CAACATAATC	ATGTTTGGCC	8850
CCACACTGTG	GGCGAGGATG	ATACTGATGA	CCCATTTCTT	TAGCGTCCTC	8900
ATAGCCAGGG	ATCAGCTTGA	ACAGGCTCTT	AAGTGTGAGA	TCTACGGAGC	8950
CTGCTACTCC	ATAGAACCAC	TGGATCTACC	TCCAATCATT	CAAAGACTCC	9000

FIG. 16E

ATGGCCTCAG	CGCATTTTCA	CTCCACAGTT	ACTCTCCAGG	TGAAATCAAT	9050
AGGGTGGCCG	CATGCCTCAG	AAAAC TTGGG	GTCCCGCCCT	TGCGAGCTTG	9100
GAGACACCCG	GCCCGGAGCG	TCCGCGCTAG	GCTTCTGTCC	AGAGGAGGCA	9150
GGGCTGCTAT	ATGTGGCAAG	TACCTCTTCA	ACTGGGCAGT	AAGAACAAAG	9200
CTCAAAC TCA	CTCCAATAGC	GGCCGCTGGC	CGGCTGGACT	TGTCCGGTTG	9250
GTTCACGGCT	GGCTACAGCG	GGGGAGACAT	TTATCACAGC	GIGTCTCATG	9300
CCCGGCCCCG	CTGGTTC TGG	TTTTCGCTAC	TCCTGCTCGC	TGCAGGGGTA	9350
GGCATCTACC	TCCTCCCCAA	CCGATGAAGG	TTGGGGTAAA	CACTCCGGCC	9400
TCTTAAGCCA	TTTCCTGTTT	TTTTTTTTTT	TTTTTTTTTT	TTTTTCTTTT	9450
TTTTTTTCTT	TCCTTTCCCT	CTTTTTTTCC	TTTCTTTTTC	CCTTCTTTAA	9500
TGGTGGCTCC	ATCTTAGCCC	TAGTCACGGC	TAGCTGTGAA	AGGTCCGTGA	9550
GCCGCATGAC	TGCAGAGAGT	GCTGATACTG	GCCTCTCTGC	AGATCATGT	9599

FIG. 16F

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
MSINPKPQRK	TKRNINRRPQ	DVKFEGGGQI	VGGVYLLPRR	GPRLGVRATR	50
KASERSQPRG	RRQPIPKARR	PEGRAWAQPG	YFWPLYGNEG	LGWAGWLLSP	100
RGSRPSWGPT	DPRRRSRNLG	KVIDTLTQGF	ADLMGYIPLV	GAPLGGAARA	150
LAHGVRVLED	GVNYATGNLP	GCSFSIFLLA	LLSCLITPAS	AYEVRNVSGI	200
YHVINDCSNS	SIVYEAADVI	MHTPGCVPCV	QEGNSSRCWV	ALTPTLAARN	250
ASVPTTTIRR	HVDLLVGTA	FCSAMYVGD	CGSIFLVSQ	FTFSPPRHET	300
VQDCNCSTYP	GHVSGHRMAW	IMMMNWSPTT	ALVVSQLLRI	PQAVVDMVAG	350
AHWGLAGLA	YYSMVGNWAK	VLIVALLFAG	VDGETHTTGR	VAGHTTSGFT	400
SLFSSGASQK	IQLVNINGSW	HINRTALN	DSLQTGFEEA	LFYAHKFNSS	450
GCPERMASCR	PIDWFAQGWG	PITYTKPNSS	DQRPYCWYHA	PRFCGVVPAS	500
QVCGFVYCFT	PSPVWGTID	RSGVPTYSWG	ENETDMMLN	NIRPPQGNWF	550
GCTWMNSTGF	TKTCGGPPCN	IGGVGNRTLI	CPTDCFRKHP	EATYTKCGSG	600
PWLTPRCLVD	YPYRLWHYPC	TLNFSIFKVR	MYVGGVEHRL	NAACNWIRGE	650
RCNLEDNRDS	ELSPLLLSTT	EWQILPCAFT	TLPALSTGLI	HLHQNTVDVQ	700
YLYGVGSAFV	SFAIKWEYIL	LLFLLLADAR	VCACLWMLL	IAQAEAALEN	750
LVVLNAASVA	GAHGILSFLV	FFCAAWYIKG	RLAPGAAYAF	YGVWPLLLLL	800
LALPPRAYAL	DTEVAASCGG	VVLVGLMALT	LSPYYKRYIS	WCMWWLQYFL	850
TRVEAQLHW	VPELNVRGGR	DAVILLMCVV	HPTLVFDITK	LLLAIFGPILW	900
ILQASLLKVP	YFVRVQGLLR	ICALARKIAG	GHYVQMAIHK	LGALTGTIVY	950
NHLTPLRDWA	HNGLRDLAVA	VEPVVFSRME	TKLITWGADT	AACGDIINGL	1000
PVSARRQGEI	LLGPADGMVS	KWRLLAPIT	AYAQQTRGLL	GCIITSLTGR	1050
DKNQVEGEVQ	IVSTATQTFL	ATCINGVCWT	VYHGAGIRTI	ASPKGPVIQM	1100
YINVDQDLVG	WPAPQGSRL	TPCTCGSSDL	YLVTRHADVI	PVRRRGDSRG	1150
SLLSPRPISY	LKGSSGGPLL	CPAGHAVGLE	RAAVCTRGVA	KAVDFIPVEN	1200
LGTIMRSPVF	TDNSSPPAVP	QSFQVAHLHA	PTGSGKSTKV	PAAYAAQGYK	1250
VLVLNPSVAA	TLGFGAYMSK	AHGVDENIRT	GVRTTTTGSP	ITYSTYGFKL	1300
ADGGCSGGAY	DIICDECHS	TDATSILGIG	TVLDQAETAG	ARLVLATAT	1350
PPGSVTVSHP	NIEEVALSTT	GEIPFYGKAI	PLEVIKGRH	LIFCHSKKKC	1400
DELAACLVAL	GINAVAYYRG	LDSVIPTSG	DVVVSTDAL	MIGFTGDFDS	1450
VIDCNTCVTQ	TVDFSLDPTF	TIETITLQD	AVSRTQRRGR	TGRGKPGIYR	1500
FVAPGERPSG	MFDSSVLCEC	YDAGCAWYEL	TPAETIVRLR	AYMNTGGLPV	1550
QQDHLEFWEG	VFTGLTHIDA	HFLSQTKQSG	ENFPYLVAYQ	ATVCARAQAP	1600
PPSWDQMWKC	LIRLKPTLHG	PTPLLYRLGA	VQNEVILTHP	ITKYIMICMS	1650
ADLEVVTSTW	VLVGGVLAAL	AAYCLSTGCV	VIVGRIVLSG	KPAIIPDREV	1700
LYQEFDEMEE	CSQHLPTYEQ	GMLAEQFKQ	KALGILLQAS	RHAEVTTPAV	1750
QTNWQKLEVF	WAKHMANFIS	GIQYLAGLST	LPGNPAIASL	MAFTAAVTSP	1800
LTTGQTLLFN	ILOGWAAQL	AAPGAATAFV	GAGLAGAAIC	SVGLGKVLVD	1850
ILAGYGAGVA	GALVAFKIMS	GEVPSTEDLV	NLLPAILSPG	ALVVGWVCAA	1900

FIG. 16G

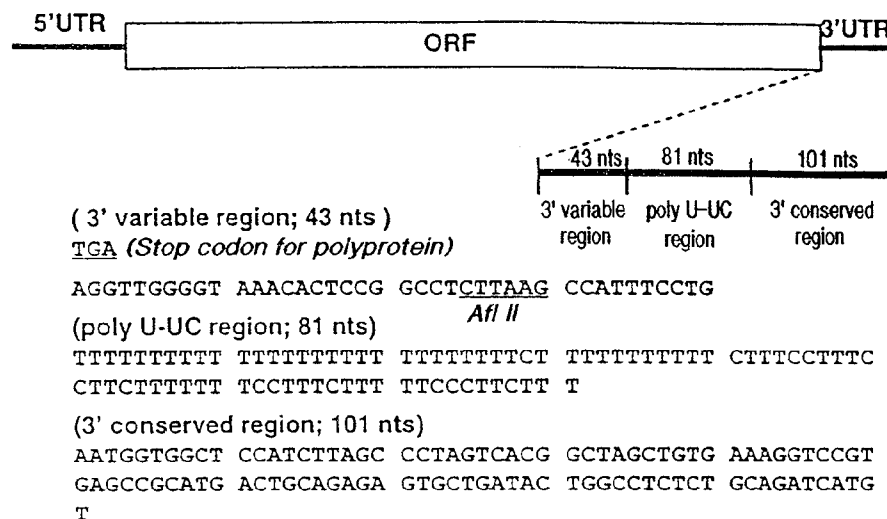
H77CV-J4aa Sequence

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ILRRHVGPGGE	GAVQWMNRLI	AFASRGNHVS	PTIHYVPESDA	AARVTAAILSS	1950
LITVTQLLRRL	HQWISSECTT	PCSGSWLRDI	WDWICEVLSD	FKTWLKAKLM	2000
PQLPGIPFVS	CQRGYRGWR	GDGIMHIRCH	CGAETTGHVK	NGIMRIVGPR	2050
TCRNWMSGTF	PINAYTTGPC	TPLPAPNYKF	ALWRVSAEEY	VEIRRVGDFH	2100
YVSGMTIDNL	KCPQIPSP	FFTELDGURL	HRFAPPCKPL	LREEVSFRVG	2150
LHEYVGSQQL	PCEPEPDVAV	LTSMLIDPSH	ITAEFAAGRRL	ARGSPPSMAS	2200
SSASQLSAPS	LKATCTANHD	SPDAELLEAN	LLWRQEMGGN	ITRVESENKV	2250
VILDSFDPLV	AEEDEREVS	PAETLRKSRR	FARALPWAR	PDYNPPLVET	2300
WKKPDYEPV	VHGCPLPPR	SPPVPPPRK	RIVVLTESTL	STALAEATK	2350
SFGSSSTSGI	TGDNITTSSE	PAPSGCPPDS	DVESYSSMPP	LEGEFGDPDL	2400
SDGSWSTVSS	GADTEDVCC	SMSYSWIGAL	VTFCAAEEQK	LPINALSNSL	2450
LRHNLVYST	TSRSACQRQK	KVTFDRLQVL	DSHYQDVLKE	VKAAASKVKA	2500
NLLSVEFACS	LTPPHSAKSK	FGYGAKDVR	HARKAVAHIN	SWKDLLED	2550
VTPIDTTIMA	KNEVFCVQPE	KGGRKPARLI	VFPDLGVRVC	EKMALYDWS	2600
KLPLAVMGSS	YGFQYSPGQR	VEFLVQAWKS	KKTFMGFSYD	TRCFDSTVTE	2650
SDIRTEEATY	QCCDLDPQAR	VAIKSLTERL	YVGGPLINSR	GENCGYRRCR	2700
ASGVLTTSCG	NILTCYIKAR	AACRAAGLQD	CTMLVCGDDL	WICESAGVQ	2750
EDAASLRAFT	EAMTRYSAAP	GDPPQPEYDL	ELITSCSSNV	SVAHGAGKR	2800
VYYLTRDPTT	PLARAAWETA	RHTPVNSWL	NIIMFAPTLW	ARMILMTHFF	2850
SVLIARDQLE	QALNCEIYGA	CYSIEPLDLP	PIIQRHLGLS	AFSLHSYSFG	2900
EINFVAACLR	KLGVPPRAW	RHRARSVRAR	LLSRGGRAAI	CGKYLENMAV	2950
RTKLKLPIA	AAGRLDLSGW	FTAGYSOGDI	YHSVSHARPR	WEWFCLLLA	3000
AGVGIIYLLN	R				3011

FIG. 16H

#1a. 3' Deletion mutants of pCV-H77C

Sequence of 3' untranslated region of pCV-H77C



#1a -1. pCV-H77C(-98X) ; 3' 98 nucleotides removed from pCV-H77C

TGAAGGTTGG GGTAACACT CCGGCCTCTT AAGCCATTTC CTGTTTTTTT
 TTTTTTTTTT TTTTTTTTTT TCTTTTTTTT TTTCTTTCCT TTCCTTCTTT
 TTTTCCTTTC TTTTCCCTT CTTTAAT

#1a -2. pCV-H77C(-42X) ; 3' 42 nucleotides removed from pCV-H77C

TGAAGGTTGG GGTAACACT CCGGCCTCTT AAGCCATTTC CTGTTTTTTT
 TTTTTTTTTT TTTTTTTTTT TCTTTTTTTT TTTCTTTCCT TTCCTTCTTT
 TTTTCCTTTC TTTTCCCTT CTTTAATGGT GGCTCCATCT TAGCCCTAGT
 CACGGCTAGC TGTGAAAGGT CCGTGAGCCG CAT

#1a -3. pCV-H77C(X-52) ; All of the 3' UTR sequence, except 3' 49 nucleotides, removed from pCV-H77C

TGAGCCGCAT GACTGCAGAG AGTGCTGATA CTGGCCTCTC TGCAGATCAT
 GT

FIG. 17A

#1a -4. pCV-H77C(X) ; All of the 3' UTR sequence, except 3' 101 nucleotides, removed from pCV-H77C

TGAAATGGTG GCTCCATCTT AGCCCTAGTC ACGGCTAGCT GTGAAAGGTC
CGTGAGCCGC ATGACTGCAG AGAGTGCTGA TACTGGCCTC TCTGCAGATC
ATGT

#1a -5. pCV-H77C(+49X) ; The proximal 49 nucleotides of the 3' conserved region (98 nucleotides ; AAT not included) removed from pCV-H77C

TGAAGGTTGG GGTAAACACT CCGGCCTCTT AAGCCATTTC CTGTTTTTTT
TTTTTTTTTT TTTTTTTTTT TCTTTTTTTT TTTCTTTCCT TTCCTTCTTT
TTTTCCCTTC TTTTCCCTT CTTTAATGCC GCATGACTGC AGAGAGTGCT
GATACTGGCC TCTCTGCAGA TCATGT

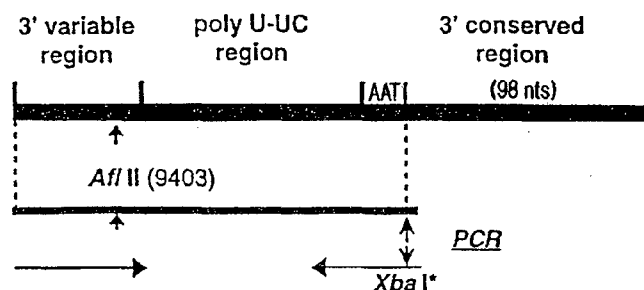
#1a -6. pCV-H77C(VR-24) ; First 24 nucleotides of the 3' variable region removed from pCV-H77C

TGACTTAAGC CATTTCTGT TTTTTTTTTT TTTTTTTTTT TTTTTTCTT
TTTTTTTTTC TTCCTTTCC TTCTTTTTTT CCTTCTTTT TCCCTTCTTT
AATGGTGGCT CCATCTTAGC CCTAGTCACG GCTAGCTGTG AAAGGTCCGT
GAGCCGCATG ACTGCAGAGA GTGCTGATAC TGGCCTCTCT GCAGATCATG
T

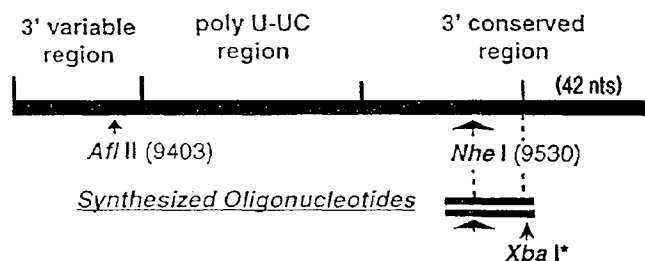
#1a -7. pCV-H77C(-U/UC) ; Poly U-UC region removed from pCV-H77C

TGAAGGTTGG GGTAAACACT CCGGCCTCTT AAGCCATTTC CTGAATGGTG
GCTCCATCTT AGCCCTAGTC ACGGCTAGCT GTGAAAGGTC CGTGAGCCGC
ATGACTGCAG AGAGTGCTGA TACTGGCCTC TCTGCAGATC ATGT

FIG. I7B

#1b. Strategy of 3' Deletion mutants**#1b -1. pCV-H77C(-98X)**

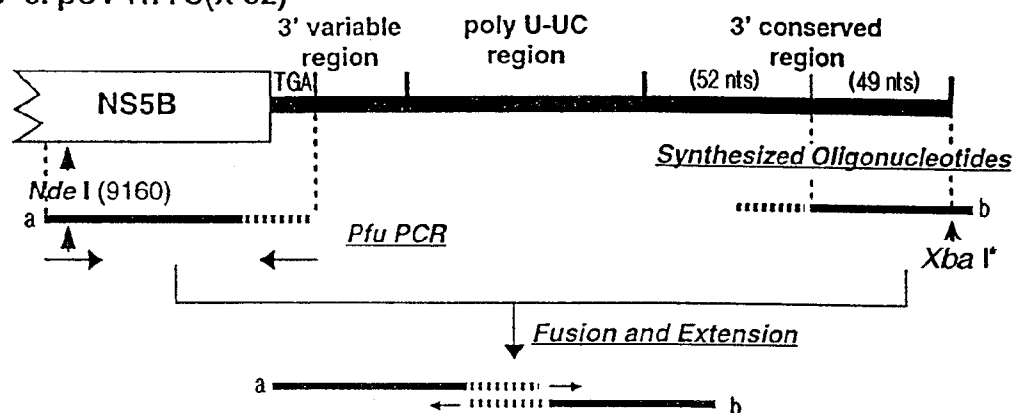
1. PCR Amplification
2. Purification of PCR products
3. Digestion with *Afl* II and *Xba* I
4. Cloning of *Afl* II /*Xba* I fragment into pCV-H77C
5. Complete sequence analysis
6. in vitro transcription (within 24 hours of inoculation)
7. Percutaneous intrahepatic transfection into chimpanzee ; 11/26/97 and 12/17/97
8. Result : Negative (No replication)

#1b -2. pCV-H77C(-42X)

1. Synthesis of oligonucleotides (sense and anti-sense)
2. Hybridization of oligonucleotides
3. Digestion with *Nhe* I and *Xba* I
4. Cloning of *Nhe* I /*Xba* I fragment into pG9-KL26 (3' UTR of H77C)
5. Sequence analysis
6. Cloning of 3' UTR (-42X) [*Afl* II /*Xba* I fragment] into pCV-H77C
7. Complete sequence analysis
8. in vitro transcription (within 24 hours of inoculation)
9. Percutaneous intrahepatic transfection into chimpanzee (Schedule; 1/22/98, 2/5/98)

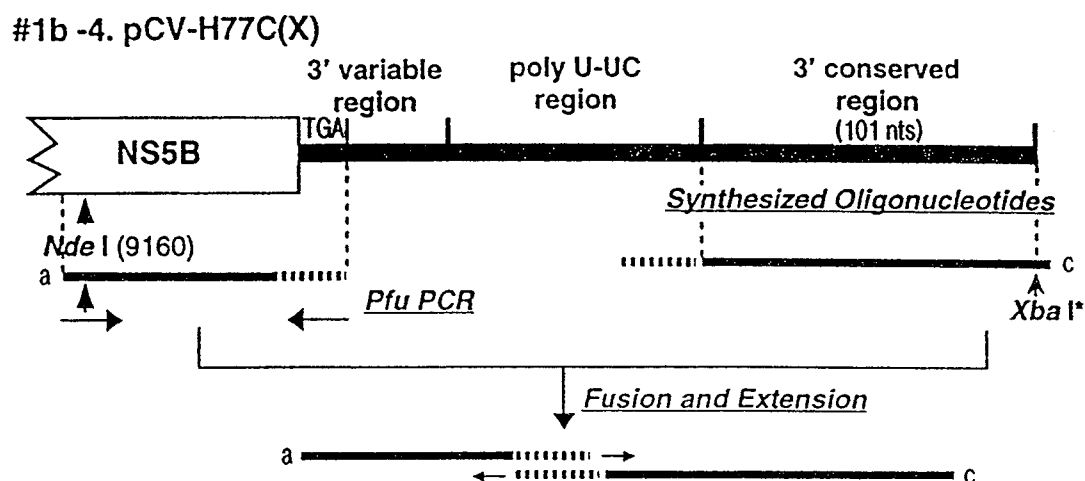
FIG. I7C

#1b -3. pCV-H77C(X-52)



1. Fragment a ; *Pfu* PCR amplification and purification
2. Fragment b ; Synthesized oligonucleotides (anti-sense)
3. Fusion and extension
4. TA cloning
5. Sequence analysis
6. Cloning *Nde*I-*Xba*I fragment with correct sequence into pCV-H77C
7. Complete sequence analysis
8. *In vitro* transcription (within 24 hours of inoculation)
9. Percutaneous intrahepatic transfection into chimpanzee

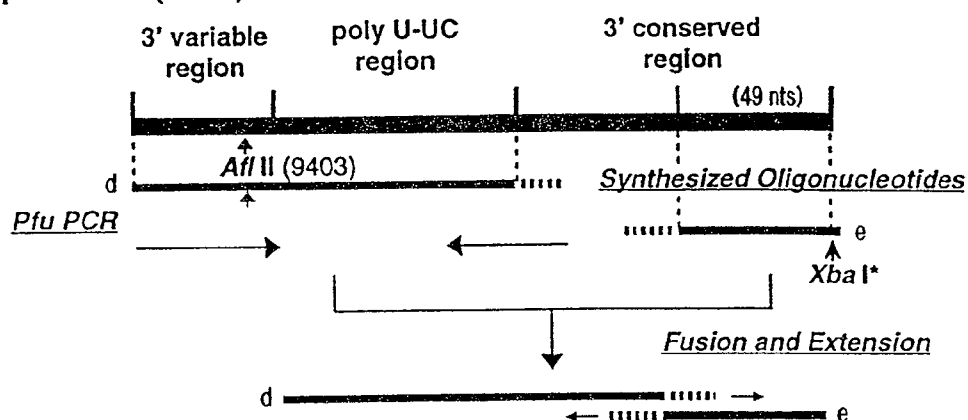
FIG. I7D



1. Fragment a ; *Pfu PCR* amplification and purification
2. Fragment c ; Synthesized oligonucleotides (anti-sense)
3. Fusion and extension
4. TA cloning
5. Sequence analysis
6. Cloning *Nde I*-*Xba I* fragment with correct sequence into pCV-H77C
7. Complete sequence analysis
8. *In vitro* transcription (within 24 hours of inoculation)
9. Percutaneous intrahepatic transfection into chimpanzee

FIG. 17E

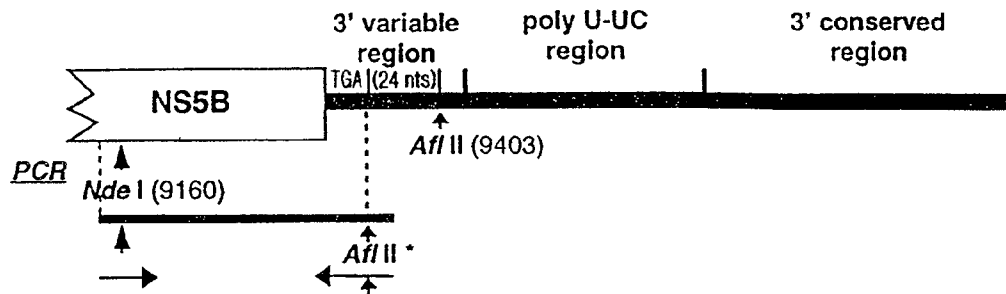
#1b -5. pCV-H77C(+49X)



1. Fragment d ; *Pfu* PCR amplification and purification
2. Fragment e ; Synthesized oligonucleotides (anti-sense)
3. Fusion and extension
4. TA cloning
5. Sequence analysis
6. Cloning *Afl* II-*Xba* I fragment with correct sequence into pCV-H77C
7. Complete sequence analysis
8. *In vitro* transcription (within 24 hours of inoculation)
9. Percutaneous intrahepatic transfection into chimpanzee

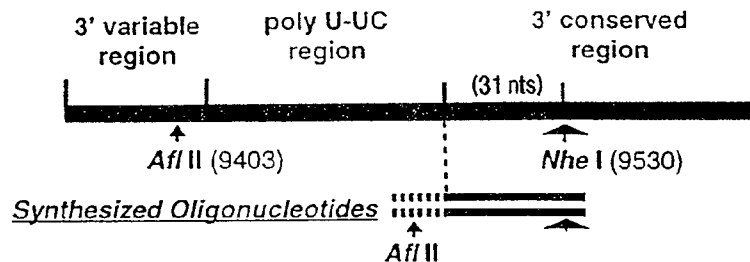
FIG. 17F

#1b -6. pCV-H77C(VR-24)



1. PCR Amplification
2. Purification of PCR products
3. Digestion with *Nde* I and *Afl* I
4. Cloning of *Nde* I / *Afl* II fragment into pCV-H77C
5. Complete sequence analysis
6. in vitro transcription (within 24 hours of inoculation)
7. Percutaneous intrahepatic transfection into chimpanzee

#1b -7. pCV-H77C(-U/UC)



1. Synthesis of oligonucleotides (sense and anti-sense)
2. Hybridization of oligonucleotides
3. Digestion with *Afl* II and *Nhe* I
4. Cloning of *Afl* II and *Nhe* I fragment into pG9-KL26
5. Sequence analysis
6. Cloning of 3' UTR (-poly U-UC) [*Afl* II / *Xba* I fragment] into pCV-H77C
7. Complete sequence analysis
8. in vitro transcription (within 24 hours of inoculation)
9. Percutaneous intrahepatic transfection into chimpanzee

FIG. 17G

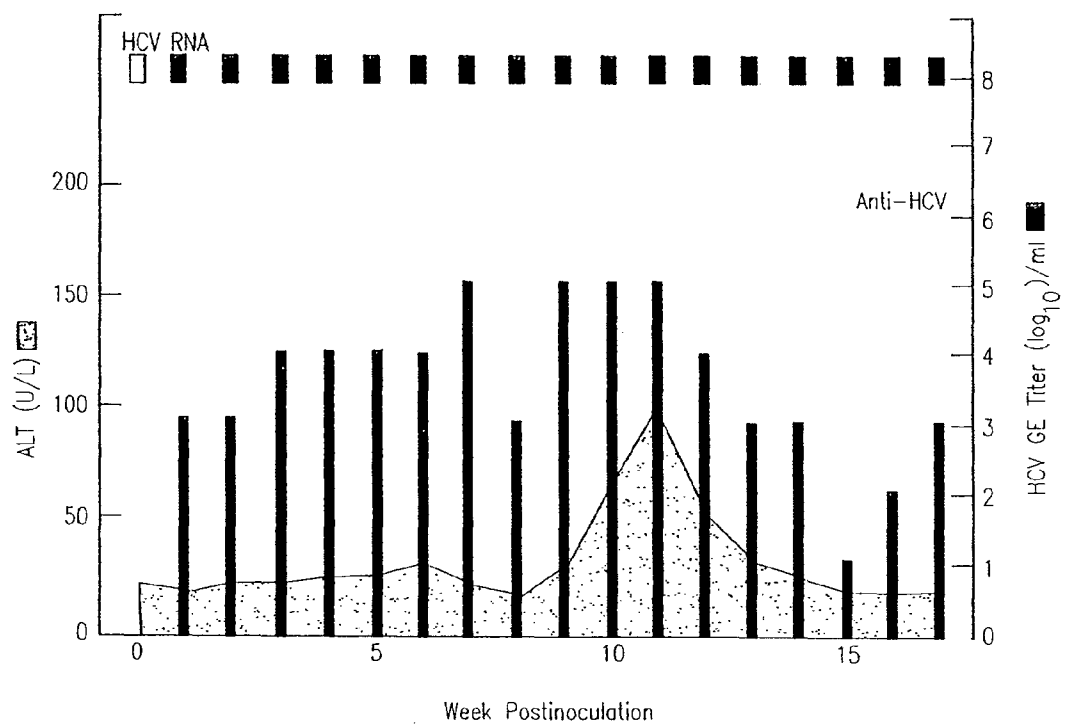


FIG. 18A

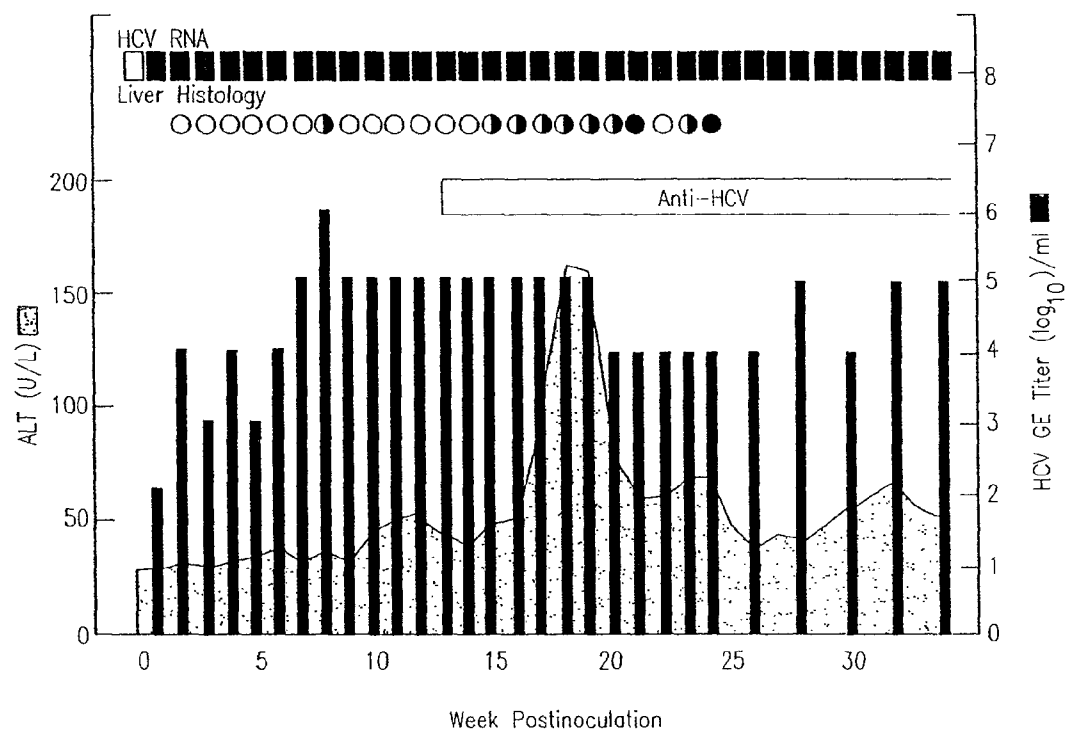


FIG. 18B

1

CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a Divisional of U.S. application Ser. No.: 09/662,454, filed Sep. 14, 2000, now U.S. Pat. No.: 7,201,911, which is a Divisional of U.S. application Ser. No.: 09/014,416, filed Jan. 27, 1998, now U.S. Pat. No.: 6,153,421, which claims the benefit of U.S. Provisional application Ser. No.: 60/053,062, filed Jul. 18, 1997, the disclosures of all of which are hereby expressly incorporated by reference in their entireties.

FIELD OF THE INVENTION

The present invention relates to molecular approaches to the production of nucleic acid sequences which comprise the genome of infectious hepatitis C viruses. In particular, the invention provides nucleic acid sequences which comprise the genomes of infectious hepatitis C viruses of genotype 1a and 1b strains. The invention therefore relates to the use of these sequences, and polypeptides encoded by all or part of these sequences, in the development of vaccines and diagnostic assays for HCV and in the development of screening assays for the identification of antiviral agents for HCV.

DESCRIPTION OF THE RELATED ART

Hepatitis C virus (HCV) has a positive-sense single-strand RNA genome and is a member of the virus family Flaviviridae (Choo, Q.-L., et al. 1991 *PNAS USA* 88:2451-2455; Rice, C. M. 1996 Flaviviridae: The viruses and their replication, In "Fields Virology". (B. N. Fields, D. M. Knipe, P. M. Howley, et al., Eds.), Third ed., Lippincott-Raven Publishers, Philadelphia). As for all positive-stranded RNA viruses, the genome of HCV functions as mRNA from which all viral proteins necessary for propagation are translated.

The viral genome of HCV is approximately 9600 nucleotides (nts) and consists of a highly conserved 5' untranslated region (UTR), a single long open reading frame (ORF) of approximately 9,000 nts and a complex 3' UTR. The 5' UTR contains an internal ribosomal entry site (Tsukiyama-Kohara, K. et al. 1992 *J. Virol.* 66:1476-1483; Honda, M. et al. 1996 *RNA* 2:955-968). The 3' UTR consists of a short variable region, a polypyrimidine tract of variable length and, at the 3' end, a highly conserved region of approximately 100 nts (Kolykhalov, A. A. et al. 1996 *J. Virol.* 70:3363-3371; Tanaka, T. et al. 1995 *Biochem. Biophys. Res. Commun.* 215:744-749; Tanaka, T. et al. 1996 *J. Virol.* 70:3307-3312; Yamada, N. et al. 1996 *Virology* 223:255-261). The last 46 nucleotides of this conserved region were predicted to form a stable stem-loop structure thought to be critical for viral replication (Blight, K. J. and Rice, C. M. 1997 *J. Virol.* 71:7345-7352; Ito, T. and Lai, M. M. C. 1997 *J. Virol.* 71:8698-8706; Tsuchihara, K. et al. 1997 *J. Virol.* 71:6720-6726). The ORF encodes a large polypeptide precursor that is cleaved into at least 10 proteins by host and viral proteinases (Rice, C. M. 1996 Flaviviridae: The viruses and their replication, In "Fields Virology". (B. N. Fields, D. M. Knipe, P. M. Howley, et al., Eds.), Third ed., Lippincott-Raven Publishers, Philadelphia). The predicted envelope proteins contain several conserved N-linked glycosylation sites and cysteine residues (Okamoto, H. et al. 1992 *Virology* 188:331-341). The NS3 gene encodes a serine protease and an RNA helicase and the NS5B gene encodes an RNA-dependent RNA polymerase.

2

Globally, six major HCV genotypes (genotypes 1-6) and multiple subtypes (a, b, c, etc.) have been identified (Bukh, J. et al. 1993 *PNAS USA* 90:8234-8238; Simmonds, P. et al. 1993 *J. Gen. Virol.* 74:2391-2399). The most divergent HCV isolates differ from each other by more than 30% over the entire genome (Okamoto, H. et al. 1992 *Virology* 188:331-341) and HCV circulates in an infected individual as a quasispecies of closely related genomes (Bukh, J. et al. 1995 *Semin. Liver Dis.* 15:41-63; Farci, P. et al. 1997 *Springer Semin. Immunopathol.* 19:5-26).

At present, more than 80% of individuals infected with HCV become chronically infected and these chronically infected individuals have a relatively high risk of developing chronic hepatitis, liver cirrhosis and hepatocellular carcinoma (Hoofnagle, J. H. 1997 *Hepatology* 26:15S-20S). In the U.S., HCV genotypes 1a and 1b constitute the majority of infections while in many other areas, especially in Europe and Japan, genotype 1b predominates.

The only effective therapy for chronic hepatitis C, interferon (IFN), induces a sustained response in less than 25% of treated patients (Fried, M. W. and Hoofnagle, J. H. 1995 *Semin. Liver Dis.* 15:82-91). Consequently, HCV is currently the most common cause of end stage liver failure and the reason for about 30% of liver transplants performed in the U.S. (Hoofnagle, J. H. 1997 *Hepatology* 26:15S-20S). In addition, a number of recent studies suggested that the severity of liver disease and the outcome of therapy may be genotype-dependent (reviewed in Bukh, J. et al. 1997 Genetic heterogeneity of hepatitis C virus and related viruses. In "Viral Hepatitis and Liver Disease, Proceedings of IX Triennial International Symposium on Viral Hepatitis and Liver Disease, Rome, Italy, 1996" (M. Rizzetto, R. H. Purcell, J. L. Gerin and G. Verme, Eds.), pp. 167-175. Edizioni Minerva Medica, Turin). In particular, these studies suggested that infection with HCV genotype 1b was associated with more severe liver disease (Brecht, C. 1997 *Hepatology* 25:772-774) and a poorer response to IFN therapy (Fried, M. W. and Hoofnagle, J. H. 1995 *Semin. Liver Dis.* 15:82-91). As a result of the inability to develop a universally effective therapy against HCV infection, it is estimated that there are still more than 25,000 new infections yearly in the U.S. (Alter, M. J. 1997 *Hepatology* 26:62S-65S). Moreover, since there is no vaccine for HCV, HCV remains a serious public health problem.

However, despite the intense interest in the development of vaccines and therapies for HCV, progress has been hindered by the absence of a useful cell culture system and the lack of any small animal model for laboratory study. For example, while replication of HCV in several cell lines has been reported, such observations have turned out not to be highly reproducible. In addition, the chimpanzee is the only animal model, other than man, for this disease. Consequently, HCV has been able to be studied only by using clinical materials obtained from patients or experimentally infected chimpanzees (an animal model whose availability is very limited).

However, several researchers have recently reported the construction of infectious cDNA clones of HCV, the identification of which would permit a more effective search for susceptible cell lines and facilitate molecular analysis of the viral genes and their function. For example, Dash, S. et al. 1997 *Am. J. Pathol.* 151: 363-373 and Yoo, B. J. et al. 1995 *J. Virol.* 69:32-38 reported that RNA transcripts from cDNA clones of HCV-1 (genotype 1a) and HCV-N (genotype 1b), respectively, resulted in viral replication after transfection into human hepatoma cell lines. Unfortunately, the viability of these clones was not tested in vivo and concerns were raised about the infectivity of these cDNA clones in vitro

(Fausto, N. 1997 *Am. J. Pathol.* 151:361). In addition, both clones did not contain the terminal 98 conserved nucleotides at the very 3' end of the UTR.

Kolykhalov, A. A., et al. 1997 *Science* 277:570-574 and Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743 reported the derivation from HCV strain H77 (which is genotype 1a) of cDNA clones of HCV that are infectious for chimpanzees. However, while these infectious clones will aid in studying HCV replication and pathogenesis and will provide an important tool for development of in vitro replication and propagation systems, it is important to have infectious clones of more than one genotype given the extensive genetic heterogeneity of HCV and the potential impact of such heterogeneity on the development of effective therapies and vaccines for HCV.

SUMMARY OF THE INVENTION

The present invention relates to nucleic acid sequences which comprise the genome of infectious hepatitis C viruses and in particular, nucleic acid sequences which comprise the genome of infectious hepatitis C viruses of genotype 1a and 1b strains. It is therefore an object of the invention to provide nucleic acid sequences which encode infectious hepatitis C viruses. Such nucleic acid sequences are referred to throughout the application as "infectious nucleic acid sequences".

For the purposes of this application, nucleic acid sequence refers to RNA, DNA, cDNA or any variant thereof capable of directing host organism synthesis of a hepatitis C virus polypeptide. It is understood that nucleic acid sequence encompasses nucleic acid sequences, which due to degeneracy, encode the same polypeptide sequence as the nucleic acid sequences described herein.

The invention also relates to the use of the infectious nucleic acid sequences to produce chimeric genomes consisting of portions of the open reading frames of infectious nucleic acid sequences of other genotypes (including, but not limited to, genotypes 1, 2, 3, 4, 5 and 6) and subtypes (including, but not limited to, subtypes 1a, 1b, 2a, 2b, 2c, 3a 4a-4f, Sa and 6a) of HCV. For example infectious nucleic acid sequence of the 1a and 1b strains H77 and HC-J4, respectively, described herein can be used to produce chimeras with sequences from the genomes of other strains of HCV from different genotypes or subtypes. Nucleic acid sequences which comprise sequence from the open-reading frames of 2 or more HCV genotypes or subtypes are designated "chimeric nucleic acid sequences".

The invention further relates to mutations of the infectious nucleic acid sequences of the invention where mutation includes, but is not limited to, point mutations, deletions and insertions. In one embodiment, a gene or fragment thereof can be deleted to determine the effect of the deleted gene or genes on the properties of the encoded virus such as its virulence and its ability to replicate. In an alternative embodiment, a mutation may be introduced into the infectious nucleic acid sequences to examine the effect of the mutation on the properties of the virus in the host cell.

The invention also relates to the introduction of mutations or deletions into the infectious nucleic acid sequences in order to produce an attenuated hepatitis C virus suitable for vaccine development.

The invention further relates to the use of the infectious nucleic acid sequences to produce attenuated viruses via passage in vitro or in vivo of the viruses produced by transfection of a host cell with the infectious nucleic acid sequence.

The present invention also relates to the use of the nucleic acid sequences of the invention or fragments thereof in the production of polypeptides where "nucleic acid sequences of

the invention" refers to infectious nucleic acid sequences, mutations of infectious nucleic acid sequences, chimeric nucleic acid sequences and sequences which comprise the genome of attenuated viruses produced from the infectious nucleic acid sequences of the invention. The polypeptides of the invention, especially structural polypeptides, can serve as immunogens in the development of vaccines or as antigens in the development of diagnostic assays for detecting the presence of HCV in biological samples.

The invention therefore also relates to vaccines for use in immunizing mammals especially humans against hepatitis C. In one embodiment, the vaccine comprises one or more polypeptides made from a nucleic acid sequence of the invention or fragment thereof. In a second embodiment, the vaccine comprises a hepatitis C virus produced by transfection of host cells with the nucleic acid sequences of the invention.

The present invention therefore relates to methods for preventing hepatitis C in a mammal. In one embodiment the method comprises administering to a mammal a polypeptide or polypeptides encoded by a nucleic acid sequence of the invention in an amount effective to induce protective immunity to hepatitis C. In another embodiment, the method of prevention comprises administering to a mammal a hepatitis C virus of the invention in an amount effective to induce protective immunity against hepatitis C.

In yet another embodiment, the method of protection comprises administering to a mammal a nucleic acid sequence of the invention or a fragment thereof in an amount effective to induce protective immunity against hepatitis C.

The invention also relates to hepatitis C viruses produced by host cells transfected with the nucleic acid sequences of the present invention.

The invention therefore also provides pharmaceutical compositions comprising the nucleic acid sequences of the invention and/or their encoded hepatitis C viruses. The invention further provides pharmaceutical compositions comprising polypeptides encoded by the nucleic acid sequences of the invention or fragments thereof. The pharmaceutical compositions of the invention may be used prophylactically or therapeutically.

The invention also relates to antibodies to the hepatitis C viruses of the invention or their encoded polypeptides and to pharmaceutical compositions comprising these antibodies.

The present invention further relates to polypeptides encoded by the nucleic acid sequences of the invention fragments thereof. In one embodiment, said polypeptide or polypeptides are fully or partially purified from hepatitis C virus produced by cells transfected with nucleic acid sequence of the invention. In another embodiment, the polypeptide or polypeptides are produced recombinantly from a fragment of the nucleic acid sequences of the invention. In yet another embodiment, the polypeptides are chemically synthesized.

The invention also relates to the use of the nucleic acid sequences of the invention to identify cell lines capable of supporting the replication of HCV in vitro.

The invention further relates to the use of the nucleic acid sequences of the invention or their encoded proteases (e.g. NS3 protease) to develop screening assays to identify antiviral agents for HCV.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows a strategy for constructing full-length cDNA clones of HCV strain H77. The long PCR products amplified with H1 and H9417R primers were cloned directly into pGEM-9zf(-) after digestion with Not I and Xba I (pH21₁ and

5

pH50₁). Next, the 3' UTR was cloned into both pH21₁, and pH50₁, after digestion with AflII and XbaI (pH21 and pH50). pH21 was tested for infectivity in a chimpanzee. To improve the efficiency of cloning, we constructed a cassette vector with consensus 5' and 3' termini of H77. This cassette vector (pCV) was obtained by cutting out the BamHI fragment (nts 1358-7530 of the H77 genome) from pH50, followed by religation. Finally, the long PCR products of H77 amplified with primers H1 and H9417R (H product) or primers A1 and H9417R (A product) were cloned into pCV after digestion with AgeI and AflII or with PinAI and BfrI. The latter procedure yielded multiple complete cDNA clones of strain H77 of HCV.

FIG. 2 shows the results of gel electrophoresis of long RT-PCR amplicons of the entire ORF of H77 and the Transcription mixture of the infectious clone of H77. The complete ORF was amplified by long RT-PCR with the primers H1 or A1 and H9417R from 10⁵ GE of H77. A total of 10 µg of the consensus chimeric clone (pCV-H77C) linearized with XbaI was transcribed in a 100 µl reaction with T7 RNA polymerase. Five µl of the transcription mixture was analyzed by gel electrophoresis and the remainder of the mixture was injected into a chimpanzee. Lane 1, molecular weight marker; lane 2, products amplified with primers H1 and H9417R; lane 3, products amplified with primers A1 and H9417R; lane 4, transcription mixture containing the RNA transcripts and linearized clone pCV-H77C (12.5 kb).

FIG. 3 is a diagram of the genome organization of HCV strain H77 and the genetic heterogeneity of individual full-length clones compared with the consensus sequence of H77. Solid lines represent aa changes. Dashed lines represent silent mutations. A * in pH21 represents a point mutation at nt 58 in the 5' UTR. In the ORF, the consensus chimeric clone pCV-H77C had 11 nt differences [at positions 1625 (C→T), 2709 (T→C), 3380 (A→G), 3710 (C→T), 3914 (G→A), 4463 (T→C), 5058 (C→T), 5834 (C→T), 6734 (T→C), 7154 (C→T), and 7202 (T→C)] and one aa change (F→L at aa 790) compared with the consensus sequence of H77. This clone was infectious. Clone pH21 and pCV-H11 had 19 nts (7 aa) and 64 nts (21 aa) differences respectively, compared with the consensus sequence of H77. These two clones were not infectious. A single point mutation in the 3' UTR at nucleotide 9406 (G→A) introduced to create an AflII cleavage site is not shown.

FIGS. 4A-4F show the complete nucleotide sequence (SEQ ID NO: 2) of a H77C clone produced according to the present invention and FIGS. 4G-4H show the amino acid sequence (SEQ ID NO: 1) encoded by the H77C clone.

FIG. 5 shows an agarose gel of long RT-PCR amplicons and transcription mixtures. Lanes 1 and 4: Molecular weight marker (Lambda/HindIII digest). Lanes 2 and 3: RT-PCR amplicons of the entire ORF of HC-J4. Lane 5: pCV-H77C transcription control (Yanagi et al., 1997). Lanes 6, 7, and 8: 1/40 of each transcription mixture of pCV-J4L2S, pCV-J4L4S and pCV-J4L6S, respectively, which was injected into the chimpanzee.

FIG. 6 shows the strategy utilized for the construction of full-length cDNA clones of HCV strain HC-J4. The long PCR products were cloned as two separate fragments (L and S) into a cassette vector (pCV) with fixed 5' and 3' termini of HCV (Yanagi et al., 1997). Full-length cDNA clones of HC-J4 were obtained by inserting the L fragment from three pCV-J4L clones into three identical pCV-J4S9 clones after digestion with PinAI (isoschizomer of AgeI) and BfrI (isoschizomer of AflII).

FIGS. 7A through 7D shows amino acid positions with a quasispecies of HC-J4 in the acute phase plasma pool

6

obtained from an experimentally infected chimpanzee. Cons-p9: consensus amino acid sequence deduced from analysis of nine L fragments and nine S fragments (see FIG. 6). Cons-D: consensus sequence derived from direct sequencing of the PCR product. A, B, C: groups of similar viral species. Dot: amino acid identical to that in Cons-p9. Capital letter: amino acid different from that in Cons-p9. Cons-F: composite consensus amino acid sequence combining Cons-p9 and Cons-D. Boxed amino acid: different from that in Cons-F. Shaded amino acid: different from that in all species A sequences. An *: defective ORF due to a nucleotide deletion (clone L1, aa 1097) or insertion (clone L7, aa 2770). Diagonal lines: fragments used to construct the infectious clone.

FIG. 8 shows comparisons (percent difference) of nucleotide (nts. 156-8935) and predicted amino acid sequences (aa 1-2864) of L clones (species A, B, and C, this study), HC-J4/91 (Okamoto, H. et al. 1992 *Virology* 190:894-899) and HC-J4/83 (Okamoto, H. et al. 1992 *Virology* 190:894-899). Differences among species A sequences and among species B sequences are shaded.

FIG. 9 shows UPGMA ("unweighted pair group method with arithmetic mean") trees of HC-J4/91 (Okamoto, H. et al. 1992 *Virology* 190:894-899), HC-J4/83 (Okamoto, H. et al. 1992 *Virology* 190:894-899), two prototype strains of genotype 1b (HCV-J, Kato, N. et al. 1990 *PNAS USA* 87:9524-9528; HCV-BK, Iakamizawa et al., 1991), and L clones (this study).

FIG. 10 shows the alignment of the HVR1 (SEQ ID NOS: 28, 30, 32, 34, 36-38, 41, 43 and 45) and HVR2 (SEQ ID NOS: 29, 31, 33, 35, 39, 40, 42, 44 and 46) amino acid sequences of the E2 sequences of nine L clones of HC-J4 (species A, B, and C) obtained from an early acute phase plasma pool of an experimentally infected chimpanzee compared with the sequences of eight clones (HC-J4/91-20 through HC-J4/91-27, Okamoto, H. et al. 1992 *Virology* 190:894-899) derived from the inoculum. Dot: an amino acid identical to that in the top line. Capital letters: amino acid different from that in the top line.

FIG. 11 shows the alignment of the 5' UTR and the 3' UTR sequences of infectious clones of genotype 1a (pCV-H77C) and 1b (pCV-J4L6S). 5' UTR for HC-J4 is SEQ ID NO: 47, 5' UTR for pCV-J4L6S is SEQ ID NO: 48, 5' UTR for pCV-H77C is SEQ ID NO: 49, 3' UTR-3' variable region for HC-J4 is SEQ ID NO: 50 and 53, 3' UTR-3' variable region for pCV-J4L6S is SEQ ID NO: 51 and 54, 3' UTR-3' variable region for pCV-H77C is SEQ ID NO: 52 and 54; 3' UTR-3' conserved region for H77, pCV-J4L6S and pCV-H77C is SEQ ID NO: 55. Top line: consensus sequence of the indicated strain. Dot: identity with consensus sequence. Capital letter: different from the consensus sequence. Dash: deletion. Underlined: PinAI and BfrI cleavage site. Numbering corresponds to the HCV sequence of pCV-J4L6S.

FIG. 12 shows a comparison of individual full-length cDNA clones of the ORF of HCV strain HC-J4 with the sequence (see FIG. 7). Solid lines: amino acid changes. Dashed lines: silent mutations. Clone pCV-J4L6S was infectious in vivo whereas clones pCV-J4L2S and pCV-J4L4S were not.

FIG. 13 shows biochemical (ALT levels) and PCR analyses of a chimpanzee following percutaneous intrahepatic transfection with RNA transcripts of the infectious clone of pCV-J4L2S, pCV-J4L4S and pCV-J4L6S. The ALT serum enzyme levels were measured in units per liter (u/l). For the PCR analysis, "HCV RNA" represented by an open rectangle indicates a serum sample that was negative for HCV after nested PCR; "HCV RNA" represented by a closed rectangle indi-

cates that the serum sample was positive for HCV and HCV GE titer on the right-hand y-axis represents genome equivalents.

FIGS. 14A-14F show the nucleotide sequence of the infectious clone of genotype 1b strain HC-J4 (SEQ ID NO: 4) and FIGS. 14G-14H show the amino acid sequence (SEQ ID NO: 3) encoded by the HC-J4 clone.

FIG. 15 shows the strategy for constructing a chimeric HCV clone designated pH77CV-J4 which contains the non-structural region of the infectious clone of genotype 1a strain H77 and the structural region of the infectious clone of genotype 1b strain HC-J4.

FIGS. 16A-16F show the nucleotide sequence of the chimeric 1a/1b clone pH77CV-J4 (SEQ ID NO: 6) of FIG. 15 and FIGS. 16G-16H show the amino acid sequence encoded by the chimeric 1a/1b clone (SEQ ID NO: 5).

FIGS. 17A and 17B show the sequence of the 3' untranslated region remaining in various 3' deletion mutants of the 1a infectious clone pCV-H77C (pCV-H77C has SEQ ID NOS: 56, 57 and 58; pCV-H77C (-98X) has SEQ ID NO: 59; pCV-H77C (-42X) has SEQ ID NO: 60; pCV-H77C (X-52) has SEQ ID NO: 61; pCV-H77C(X) has SEQ ID NO: 62; pCV-H77C(+49X) has SEQ ID NO: 63; pCV-H77C (VR-24) has SEQ ID NO: 64; and pCV-H77C (-U/UC) has SEQ ID NO: 65) and the strategy utilized in constructing each 3' deletion mutant (FIGS. 17C-17G).

Of the seven deletion mutants shown, two (pCV-H77C(-98X) and (-42X)) have been constructed and tested for infectivity in chimpanzees (see FIGS. 17A and 17C) and the other six are to be constructed and tested for infectivity as described in FIGS. 17D-17G.

FIGS. 18A and 18B show biochemical (ALT levels), PCR (HCV RNA and HCV GE titer), serological (anti-HCV) and histopathological (FIG. 18B only) analyses of chimpanzees 1494 (FIG. 18A) and 1530 (FIG. 18B) following transfection with the infectious cDNA clone pCV-H77C.

The AL-T serum enzyme levels were measured in units per ml (u/i). For the PCR analysis, "HCV RNA" represented by an open rectangle indicates a serum sample that was negative for HCV after nested PCR; "HCV RNA" represented by a closed rectangle indicates that the serum sample was positive for HCV; and HCV GE titer on the right-hand y-axis represents genome equivalents.

The bar marked "anti-HCV" indicates samples that were positive for anti-HCV antibodies as determined by commercial assays. The histopathology scores in FIG. 18B correspond to no histopathology (○), mild hepatitis (⊖) and moderate to severe hepatitis (●).

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to nucleic acid sequences which comprise the genome of an infectious hepatitis C virus. More specifically, the invention relates to nucleic acid sequences which encode infectious hepatitis C viruses of genotype 1a and 1b strains. In one embodiment, the infectious nucleic acid sequence of the invention has the sequence shown in FIGS. 4A-4F of this application. In another embodiment, the infectious nucleic acid sequence has the sequence shown in FIGS. 14A-14F and is contained in a plasmid construct deposited with the American Type Culture Collection (ATCC) on Jan. 26, 1998 and having ATCC accession number 209596.

The invention also relates to "chimeric nucleic acid sequences" where the chimeric nucleic acid sequences con-

sist of open-reading frame sequences taken from infectious nucleic acid sequences of hepatitis C viruses of different genotypes or subtypes.

In one embodiment, the chimeric nucleic acid sequence consists of sequence from the genome of an HCV strain belonging to one genotype or subtype which encodes structural polypeptides and sequence of an HCV strain belonging to another genotype strain or subtype which encodes non-structural polypeptides. Such chimeras can be produced by standard techniques of restriction digestion, PCR amplification and subcloning known to those of ordinary skill in the art.

In a preferred embodiment, the sequence encoding non-structural polypeptides is from an infectious nucleic acid sequence encoding a genotype 1a strain where the construction of a chimeric 1a/1b nucleic acid sequence is described in Example 9 and the chimeric 1a/1b nucleic acid sequence is shown in FIGS. 16A-16F. It is believed that the construction of such chimeric nucleic acid sequences will be of importance in studying the growth and virulence properties of hepatitis C virus and in the production of hepatitis C viruses suitable to confer protection against multiple genotypes of HCV. For example, one might produce a "multivalent" vaccine by putting epitopes from several genotypes or subtypes into one clone. Alternatively one might replace just a single gene from an infectious sequence with the corresponding gene from the genomic sequence of a strain from another genotype or subtype or create a chimeric gene which contains portions of a gene from two genotypes or subtypes. Examples of genes which could be replaced or which could be made chimeric, include, but are not limited to, the E1, E2 and NS4 genes.

The invention further relates to mutations of the infectious nucleic acid sequences where "mutations" includes, but is not limited to, point mutations, deletions and insertions. Of course, one of ordinary skill in the art would recognize that the size of the insertions would be limited by the ability of the resultant nucleic acid sequence to be properly packaged within the virion. Such mutation could be produced by techniques known to those of skill in the art such as site-directed mutagenesis, fusion PCR, and restriction digestion followed by religation.

In one embodiment, mutagenesis might be undertaken to determine sequences that are important for viral properties such as replication or virulence. For example, one may introduce a mutation into the infectious nucleic acid sequence which eliminates the cleavage site between the NS4A and NS4B polypeptides to examine the effects on viral replication and processing of the polypeptide. Alternatively, one or more of the 3 amino acids encoded by the infectious 1b nucleic acid sequence shown in FIGS. 14A-14F which differ from the HC-J4 consensus sequence may be back mutated to the corresponding amino acid in the HC-J4 consensus sequence to determine the importance of these three amino acid changes to infectivity or virulence. In yet another embodiment, one or more of the amino acids from the noninfectious 1b clones pCV-J4L2S and pCV-J4L4S which differ from the consensus sequence may be introduced into the infectious 1b sequence shown in FIGS. 14A-14F.

In yet another example, one may delete all or part of a gene or of the 5' or 3' nontranslated region contained in an infectious nucleic acid sequence and then transfect a host cell (animal or cell culture) with the mutated sequence and measure viral replication in the host by methods known in the art such as RT-PCR. Preferred genes include, but are not limited to, the P7, NS4B and NS5A genes. Of course, those of ordinary skill in the art will understand that deletion of part of a gene, preferably the central portion of the gene, may be preferable to deletion of the entire gene in order to conserve the

cleavage site boundaries which exist between proteins in the HCV polyprotein and which are necessary for proper processing of the polyprotein.

In the alternative, if the transfection is into a host animal such as a chimpanzee, one can monitor the virulence phenotype of the virus produced by transfection of the mutated infectious nucleic acid sequence by methods known in the art such as measurement of liver enzyme levels (alanine aminotransferase (ALT) or isocitrate dehydrogenase (ICD)) or by histopathology of liver biopsies. Thus, mutations of the infectious nucleic acid sequences may be useful in the production of attenuated HCV strains suitable for vaccine use.

The invention also relates to the use of the infectious nucleic acid sequences of the present invention to produce attenuated viral strains via passage in vitro or in vivo of the virus produced by transfection with the infectious nucleic acid sequences.

The present invention therefore relates to the use of the nucleic acid sequences of the invention to identify cell lines capable of supporting the replication of HCV.

In particular, it is contemplated that the mutations of the infectious nucleic acid sequences of the invention and the production of chimeric sequences as discussed above may be useful in identifying sequences critical for cell culture adaptation of HCV and hence, may be useful in identifying cell lines capable of supporting HCV replication.

Transfection of tissue culture cells with the nucleic acid sequences of the invention may be done by methods of transfection known in the art such as electroporation, precipitation with DEAE-Dextran or calcium phosphate or liposomes.

In one such embodiment, the method comprises the growing of animal cells, especially human cells, in vitro and transfecting the cells with the nucleic acid of the invention, then determining if the cells show indicia of HCV infection. Such indicia include the detection of viral antigens in the cell, for example, by immunofluorescent procedures well known in the art; the detection of viral polypeptides by Western blotting using antibodies specific therefor; and the detection of newly transcribed viral RNA within the cells via methods such as RT-PCR. The presence of live, infectious virus particles following such tests may also be shown by injection of cell culture medium or cell lysates into healthy, susceptible animals, with subsequent exhibition of the symptoms of HCV infection.

Suitable cells or cell lines for culturing HCV include, but are not limited to, lymphocyte and hepatocyte cell lines known in the art.

Alternatively, primary hepatocytes can be cultured, and then infected with HCV; or, the hepatocyte cultures could be derived from the livers of infected chimpanzees. In addition, various immortalization methods known to those of ordinary skill in the art can be used to obtain cell-lines derived from hepatocyte cultures. For example, primary hepatocyte cultures may be fused to a variety of cells to maintain stability.

The present invention further relates to the in vitro and in vivo production of hepatitis C viruses from the nucleic acid sequences of the invention.

In one embodiment, the sequences of the invention can be inserted into an expression vector that functions in eukaryotic cells. Eukaryotic expression vectors suitable for producing high efficiency gene transfer in vivo are well known to those of ordinary skill in the art and include, but are not limited to, plasmids, vaccinia viruses, retroviruses, adenoviruses and adeno-associated viruses.

In another embodiment, the sequences contained in the recombinant expression vector can be transcribed in vitro by methods known to those of ordinary skill in the art in order to

produce RNA transcripts which encode the hepatitis C viruses of the invention. The hepatitis C viruses of the invention may then be produced by transfecting cells by methods known to those of ordinary skill in the art with either the in vitro transcription mixture containing the RNA transcripts (see Example 4) or with the recombinant expression vectors containing the nucleic acid sequences described herein.

The present invention also relates to the construction of cassette vectors useful in the cloning of viral genomes wherein said vectors comprise a nucleic acid sequence to be cloned, and said vector reading in the correct phase for the expression of the viral nucleic acid to be cloned. Such a cassette vector will, of course, also possess a promoter sequence, advantageously placed upstream of the sequence to be expressed. Cassette vectors according to the present invention are constructed according to the procedure described in FIG. 1, for example, starting with plasmid pCV. Of course, the DNA to be inserted into said cassette vector can be derived from any virus, advantageously from HCV, and most advantageously from the H77 strain of HCV. The nucleic acid to be inserted according to the present invention can, for example, contain one or more open reading frames of the virus, for example, HCV. The cassette vectors of the present invention may also contain, optionally, one or more expressible marker genes for expression as an indication of successful transfection and expression of the nucleic acid sequences of the vector. To insure expression, the cassette vectors of the present invention will contain a promoter sequence for binding of the appropriate cellular RNA polymerase, which will depend on the cell into which the vector has been introduced. For example, if the host cell is a bacterial cell, then said promoter will be a bacterial promoter sequence to which the bacterial RNA polymerases will bind.

The hepatitis C viruses produced from the sequences of the invention may be purified or partially purified from the transfected cells by methods known to those of ordinary skill in the art. In a preferred embodiment, the viruses are partially purified prior to their use as immunogens in the pharmaceutical compositions and vaccines of the present invention.

The present invention therefore relates to the use of the hepatitis C viruses produced from the nucleic acid sequences of the invention as immunogens in live or killed (e.g., formalin inactivated) vaccines to prevent hepatitis C in a mammal.

In an alternative embodiment, the immunogen of the present invention may be an infectious nucleic acid sequence, a chimeric nucleic acid sequence, or a mutated infectious nucleic acid sequence which encodes a hepatitis C virus. Where the sequence is a cDNA sequence, the cDNAs and their RNA transcripts may be used to transfect a mammal by direct injection into the liver tissue of the mammal as described in the Examples.

Alternatively, direct gene transfer may be accomplished via administration of a eukaryotic expression vector containing a nucleic acid sequence of the invention.

In yet another embodiment, the immunogen may be a polypeptide encoded by the nucleic acid sequences of the invention. The present invention therefore also relates to polypeptides produced from the nucleic acid sequences of the invention or fragments thereof. In one embodiment, polypeptides of the present invention can be recombinantly produced by synthesis from the nucleic acid sequences of the invention or isolated fragments thereof, and purified, or partially purified, from transfected cells using methods already known in the art. In an alternative embodiment, the polypeptides may be purified or partially purified from viral particles produced via transfection of a host cell with the nucleic acid sequences of the invention. Such polypeptides might, for example,

include either capsid or envelope polypeptides prepared from the sequences of the present invention.

When used as immunogens, the nucleic acid sequences of the invention, or the polypeptides or viruses produced therefrom, are preferably partially purified prior to use as immunogens in pharmaceutical compositions and vaccines of the present invention. When used as a vaccine, the sequences and the polypeptide and virus products thereof, can be administered alone or in a suitable diluent, including, but not limited to, water, saline, or some type of buffered medium. The vaccine according to the present invention may be administered to an animal, especially a mammal, and most especially a human, by a variety of routes, including, but not limited to, intradermally, intramuscularly, subcutaneously, or in any combination thereof.

Suitable amounts of material to administer for prophylactic and therapeutic purposes will vary depending on the route selected and the immunogen (nucleic acid, virus, polypeptide) administered. One skilled in the art will appreciate that the amounts to be administered for any particular treatment protocol can be readily determined without undue experimentation. The vaccines of the present invention may be administered once or periodically until a suitable titer of anti-HCV antibodies appear in the blood. For an immunogen consisting of a nucleic acid sequence, a suitable amount of nucleic acid sequence to be used for prophylactic purposes might be expected to fall in the range of from about 100 µg to about 5 mg and most preferably in the range of from about 500 µg to about 2 mg. For a polypeptide, a suitable amount to use for prophylactic purposes is preferably 100 ng to 100 µg and for a virus 10² to 10⁶ infectious doses. Such administration will, of course, occur prior to any sign of HCV infection.

A vaccine of the present invention may be employed in such forms as capsules, liquid solutions, suspensions or elixirs for oral administration, or sterile liquid forms such as solutions or suspensions. Any inert carrier is preferably used, such as saline or phosphate-buffered saline, or any such carrier in which the HCV of the present invention can be suitably suspended. The vaccines may be in the form of single dose preparations or in multi-dose flasks which can be utilized for mass-vaccination programs of both animals and humans. For purposes of using the vaccines of the present invention reference is made to Remington's Pharmaceutical Sciences, Mack Publishing Co., Easton, Pa., Osol (Ed.) (1980); and New Trends and Developments in Vaccines, Voller et al. (Eds.), University Park Press, Baltimore, Md. (1978), both of which provide much useful information for preparing and using vaccines. Of course, the polypeptides of the present invention, when used as vaccines, can include, as part of the composition or emulsion, a suitable adjuvant, such as alum (or aluminum hydroxide) when humans are to be vaccinated, to further stimulate production of antibodies by immune cells. When nucleic acids or viruses are used for vaccination purposes, other specific adjuvants such as CpG motifs (Krieg, A. M. et al. 1995 *Nature* 374:546 and Krieg, A. M. et al. 1996 *J. Lab. Clin. Med.* 128:128), may prove useful.

When the nucleic acids, viruses and polypeptides of the present invention are used as vaccines or inocula, they will normally exist as physically discrete units suitable as a unitary dosage for animals, especially mammals, and most especially humans, wherein each unit will contain a predetermined quantity of active material calculated to produce the desired immunogenic effect in association with the required diluent. The dose of said vaccine or inoculum according to the present invention is administered at least once. In order to increase the antibody level, a second or booster dose may be administered at some time after the initial dose. The need for,

and timing of, such booster dose will, of course, be determined within the sound judgment of the administrator of such vaccine or inoculum and according to sound principles well known in the art. For example, such booster dose could reasonably be expected to be advantageous at some time between about 2 weeks to about 6 months following the initial vaccination. Subsequent doses may be administered as indicated.

The nucleic acid sequences, viruses and polypeptides of the present invention can also be administered for purposes of therapy, where a mammal, especially a primate, and most especially a human, is already infected, as shown by well known diagnostic measures. When the nucleic acid sequences, viruses or polypeptides of the present invention are used for such therapeutic purposes, much of the same criteria will apply as when it is used as a vaccine, except that inoculation will occur post-infection. Thus, when the nucleic acid sequences, viruses or polypeptides of the present invention are used as therapeutic agents in the treatment of infection, the therapeutic agent comprises a pharmaceutical composition containing a sufficient amount of said nucleic acid sequences, viruses or polypeptides so as to elicit a therapeutically effective response in the organism to be treated. Of course, the amount of pharmaceutical composition to be administered will, as for vaccines, vary depending on the immunogen contained therein (nucleic acid, polypeptide, virus) and on the route of administration.

The therapeutic agent according to the present invention can thus be administered by, subcutaneous, intramuscular or intradermal routes. One skilled in the art will certainly appreciate that the amounts to be administered for any particular treatment protocol can be readily determined without undue experimentation. Of course, the actual amounts will vary depending on the route of administration as well as the sex, age, and clinical status of the subject which, in the case of human patients, is to be determined with the sound judgment of the clinician.

The therapeutic agent of the present invention can be employed in such forms as capsules, liquid solutions, suspensions or elixirs, or sterile liquid forms such as solutions or suspensions. Any inert carrier is preferably used, such as saline, phosphate-buffered saline, or any such carrier in which the HCV of the present invention can be suitably suspended. The therapeutic agents may be in the form of single dose preparations or in the multi-dose flasks which can be utilized for mass-treatment programs of both animals and humans. Of course, when the nucleic acid sequences, viruses or polypeptides of the present invention are used as therapeutic agents they may be administered as a single dose or as a series of doses, depending on the situation as determined by the person conducting the treatment.

The nucleic acids, polypeptides and viruses of the present invention can also be utilized in the production of antibodies against HCV. The term "antibody" is herein used to refer to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules. Examples of antibody molecules are intact immunoglobulin molecules, substantially intact immunoglobulin molecules and portions of an immunoglobulin molecule, including those portions known in the art as Fab, F(ab')₂ and F(v) as well as chimeric antibody molecules.

Thus, the polypeptides, viruses and nucleic acid sequences of the present invention can be used in the generation of antibodies that immunoreact (i.e., specific binding between an antigenic determinant-containing molecule and a molecule containing an antibody combining site such as a whole

13

antibody molecule or an active portion thereof) with antigenic determinants on the surface of hepatitis C virus particles.

The present invention therefore also relates to antibodies produced following immunization with the nucleic acid sequences, viruses or polypeptides of the present invention. These antibodies are typically produced by immunizing a mammal with an immunogen or vaccine to induce antibody molecules having immunospecificity for polypeptides or viruses produced in response to infection with the nucleic acid sequences of the present invention. When used in generating such antibodies, the nucleic acid sequences, viruses, or polypeptides of the present invention may be linked to some type of carrier molecule. The resulting antibody molecules are then collected from said mammal. Antibodies produced according to the present invention have the unique advantage of being generated in response to authentic, functional, polypeptides produced according to the actual cloned HCV genome.

The antibody molecules of the present invention may be polyclonal or monoclonal. Monoclonal antibodies are readily produced by methods well known in the art. Portions of immunoglobulin molecules, such as Fabs, as well as chimeric antibodies, may also be produced by methods well known to those of ordinary skill in the art of generating such antibodies.

The antibodies according to the present invention may also be contained in blood plasma, serum, hybridoma supernatants, and the like. Alternatively, the antibody of the present invention is isolated to the extent desired by well known techniques such as, for example, using DEAE Sephadex. The antibodies produced according to the present invention may be further purified so as to obtain specific classes or subclasses of antibody such as IgM, IgG, IgA, and the like. Antibodies of the IgG class are preferred for purposes of passive protection.

The antibodies of the present invention are useful in the prevention and treatment of diseases caused by hepatitis C virus in animals, especially mammals, and most especially humans.

In providing the antibodies of the present invention to a recipient mammal, preferably a human, the dosage of administered antibodies will vary depending on such factors as the mammal's age, weight, height, sex, general medical condition, previous medical history, and the like.

In general, it will be advantageous to provide the recipient mammal with a dosage of antibodies in the range of from about 1 mg/kg body weight to about 10 mg/kg body weight of the mammal, although a lower or higher dose may be administered if found desirable. Such antibodies will normally be administered by intravenous or intramuscular route as an inoculum. The antibodies of the present invention are intended to be provided to the recipient subject in an amount sufficient to prevent, lessen or attenuate the severity, extent or duration of any existing infection.

The antibodies prepared by use of the nucleic acid sequences, viruses or polypeptides of the present invention are also highly useful for diagnostic purposes. For example, the antibodies can be used as in vitro diagnostic agents to test for the presence of HCV in biological samples taken from animals, especially humans. Such assays include, but are not limited to, radioimmunoassays, EIA, fluorescence, Western blot analysis and ELISAs. In one such embodiment, the biological sample is contacted with antibodies of the present invention and a labeled second antibody is used to detect the presence of HCV to which the antibodies are bound.

Such assays may be, for example, a direct protocol (where the labeled first antibody is immunoreactive with the antigen, such as, for example, a polypeptide on the surface of the

14

virus), an indirect protocol (where a labeled second antibody is reactive with the first antibody), a competitive protocol (such as would involve the addition of a labeled antigen), or a sandwich protocol (where both labeled and unlabeled antibody are used), as well as other protocols well known and described in the art.

In one embodiment, an immunoassay method would utilize an antibody specific for HCV envelope determinants and would further comprise the steps of contacting a biological sample with the HCV-specific antibody and then detecting the presence of HCV material in the test sample using one of the types of assay protocols as described above. Polypeptides and antibodies produced according to the present invention may also be supplied in the form of a kit, either present in vials as purified material, or present in compositions and suspended in suitable diluents as previously described.

In a preferred embodiment, such a diagnostic test kit for detection of HCV antigens in a test sample comprises in combination a series of containers, each container a reagent needed for such assay. Thus, one such container would contain a specific amount of HCV-specific antibody as already described, a second container would contain a diluent for suspension of the sample to be tested, a third container would contain a positive control and an additional container would contain a negative control. An additional container could contain a blank.

For all prophylactic, therapeutic and diagnostic uses, the antibodies of the invention and other reagents, plus appropriate devices and accessories, may be provided in the form of a kit so as to facilitate ready availability and ease of use.

The present invention also relates to the use of nucleic acid sequences and polypeptides of the present invention to screen potential antiviral agents for antiviral activity against HCV. Such screening methods are known by those of skill in the art. Generally, the antiviral agents are tested at a variety of concentrations, for their effect on preventing viral replication in cell culture systems which support viral replication, and then for an inhibition of infectivity or of viral pathogenicity (and a low level of toxicity) in an animal model system.

In one embodiment, animal cells (especially human cells) transfected with the nucleic acid sequences of the invention are cultured in vitro and the cells are treated with a candidate antiviral agent (a chemical, peptide etc.) for antiviral activity by adding the candidate agent to the medium. The treated cells are then exposed, possibly under transfecting or fusing conditions known in the art, to the nucleic acid sequences of the present invention. A sufficient period of time would then be allowed to pass for infection to occur, following which the presence or absence of viral replication would be determined versus untreated control cells by methods known to those of ordinary skill in the art. Such methods include, but are not limited to, the detection of viral antigens in the cell, for example, by immunofluorescent procedures well known in the art; the detection of viral polypeptides by Western blotting using antibodies specific therefor; the detection of newly transcribed viral RNA within the cells by RT-PCR; and the detection of the presence of live, infectious virus particles by injection of cell culture medium or cell lysates into healthy, susceptible animals, with subsequent exhibition of the symptoms of HCV infection. A comparison of results obtained for control cells (treated only with nucleic acid sequence) with those obtained for treated cells (nucleic acid sequence and antiviral agent) would indicate, the degree, if any, of antiviral activity of the candidate antiviral agent. Of course, one of ordinary skill in the art would readily understand that such cells can be treated with the candidate antiviral agent either before or after exposure to the nucleic acid sequence of the

present invention so as to determine what stage, or stages, of viral infection and replication said agent is effective against.

In an alternative embodiment, a protease such as NS3 protease produced from a nucleic acid sequence of the invention may be used to screen for protease inhibitors which may act as antiviral agents. The structural and nonstructural regions of the HCV genome, including nucleotide and amino acid locations, have been determined, for example, as depicted in Houghton, M. 1996 Hepatitis C viruses. In "Fields Virology" (B. N. Fields, D. M. Knipe, P. M. Howley, et al., Eds.), Third ed., Lippincott-Raven Publishers, Philadelphia, FIG. 1; and Major, M. E. and Feinstone, S. M. 1997 *Heptology* 25:1527-1538, Table 1.

Such above-mentioned protease inhibitors may take the form of chemical compounds or peptides which mimic the known cleavage sites of the protease and may be screened using methods known to those of skill in the art (Houghton, M. 1996 Hepatitis C viruses. In "Fields Virology" (B. N. Fields, D. M. Knipe, P. M. Howley, et al., Eds.), Third ed., Lippincott-Raven Publishers, Philadelphia and Major, M. E. and Feinstone, S. M. 1997 *Heptology* 25:1527-1538). For example, a substrate may be employed which mimics the protease's natural substrate, but which provides a detectable signal (e.g. by fluorimetric or calorimetric methods) when cleaved. This substrate is then incubated with the protease and the candidate protease inhibitor under conditions of suitable pH, temperature etc. to detect protease activity. The proteolytic activities of the protease in the presence or absence of the candidate inhibitor are then determined.

In yet another embodiment, a candidate antiviral agent (such as a protease inhibitor) may be directly assayed in vivo for antiviral activity by administering the candidate antiviral agent to a chimpanzee transfected with a nucleic acid sequence of the invention and then measuring viral replication in vivo via methods such as RT-PCR. Of course, the chimpanzee may be treated with the candidate agent either before or after transfection with the infectious nucleic acid sequence so as to determine what stage, or stages, of viral infection and replication the agent is effective against.

The invention also provides that the nucleic acid sequences, viruses and polypeptides of the invention may be supplied in the form of a kit, alone or in the form of a pharmaceutical composition.

All scientific publication and/or patents cited herein are specifically incorporated by reference. The following examples illustrate various aspects of the invention but are in no way intended to limit the scope thereof.

EXAMPLES

Materials and Methods for Examples 1-4

Collection of Virus

Hepatitis C virus was collected and used as a source for the RNA used in generating the cDNA clones according to the present invention. Plasma containing strain H77 of HCV was obtained from a patient in the acute phase of transfusion-associated non-A, non-B hepatitis (Feinstone, S. M. et al 1981 *J. Infect. Dis.* 144:588-598). Strain H77 belongs to genotype 1a of HCV (Ogata, N. et al. 1991 *PNAS USA* 88:3392-3396; Inchauspe, G. et al. 1991 *PNAS USA* 88:10292-10296). The consensus sequence for most of its genome has been determined (Kolyakov et al. (1996), Ogata, N. et al. 1991 *PNAS USA* 88:3392-3396; Inchauspe, G. et al. 1991 *PNAS USA* 88:10292-10296 and Farci, P. et al. 1996 *PNAS USA* 93:15394-15399).

RNA Purification

Viral RNA was collected and purified by conventional means. In general, total RNA from 10 µl of H77 plasma was extracted with the TRIzol system (GIBCO BRL). The RNA pellet was resuspended in 100 µl of 10 mM dithiothreitol (DTT) with 5% (vol/vol) RNasin (20-40 units/µl) (available from Promega) and 10 µl aliquots were stored at -80° C. In subsequent experiments RT-PCR was performed on RNA equivalent to 1 µl of H77 plasma, which contained an estimated 10⁶ genome equivalents (GE) of HCV (Yanagi, M. et al. 1996 *J. Infect. Dis.* 174:1324-1327).

Primers used in the RT-PCR process were deduced from the genomic sequences of strain H77 according to procedures already known in the art (see above) or else were determined specifically for use herein. The primers generated for this purpose are listed in Table 1.

TABLE 1

Oligonucleotides used for PCR amplification of strain H77 of HCV		
Designation	Sequence (5'→3')*	SEQ ID NO
H9261F	GGCTACAGCGGGGGAGACATT TATCAGC	(SEQ ID NO: 7)
H3'X58R	TCATGCGGCTCAGGACCTTTC ACAGCTAG	(SEQ ID NO: 8)
H9282F	GTCC AAGCTT ATCAGCGTGT CTCATGCCCGGCCCG	(SEQ ID NO: 9)
H3'X45R	CGTCT CTAGAG GGACCTTTCACA GCTAGCCGTGACTAGGG	(SEQ ID NO: 10)
H9375F	TGAAGTTGGGGTAAACACTCC GGCCTCTTAGGCCATT	(SEQ ID NO: 11)
H3'X-35R	ACATGATCTGCAGAGAGGCCAG TATCAGCACTCTC	(SEQ ID NO: 12)
H9386F	GTCC AAGCTTACGCGT AAACAC TCCGGCCTCCTTAAGCCATTCC TG	(SEQ ID NO: 13)
H3'X-38R	CGTCT CTAGAC ATGATCTGCAG AGAGGCCAGTATCAGCACTCTC TGC	(SEQ ID NO: 14)
H1	TTTTTTT TCGGCCGC <i>TAATAC</i> <i>GACTCACTATAG</i> CCAGCCCCCT GAT-GGGGGCGACACTCCACCA TG	(SEQ ID NO: 15)
A1	ACTGTCTTCACGCAGAAAGCGT CTAGCCAT	(SEQ ID NO: 16)
H9417R	CGTCTCTAGACAGGAAATGGCT TAAGAGGCCGGAGTGT TACC	(SEQ ID NO: 17)

*HCV sequences are shown in plain text, nonHCV-specific sequences are shown in boldface and artificial cleavage sites used for cDNA cloning are underlined. The core sequence of the T7 promoter in primer H1 is shown in italics.

Primers for long RT-PCR were size-purified.

cDNA Synthesis

The RNA was denatured at 65° C. for 2 min, and cDNA synthesis was performed in a 20 µl reaction volume with Superscript II reverse transcriptase (from GIBCO/BRL) at 42° C. for 1 hour using specific antisense primers as described previously (Tellier, R. et al. 1996 *PNAS USA* 93:4370-4373). The cDNA mixture was treated with RNase H and RNase T1 (GIBCO/BRL) for 20 min at 37° C.

Amplification and Cloning of the 3' UTR

The 3' UTR of strain H77 was amplified by PCR in two different assays. In both of these nested PCR reactions the first round of PCR was performed in a total volume of 50 μ l in 1 $\times$ buffer, 250 μ mol of each deoxynucleoside triphosphate (dNTP; Pharmacia), 20 μ mol each of external sense and antisense primers, 1 μ l of the Advantage KlenTaq polymerase mix (from Clontech) and 2 μ l of the final cDNA reaction mixture. In the second round of PCR, 5 μ l of the first round PCR mixture was added to 45 μ l of PCR mixture prepared as already described. Each round of PCR (35 cycles), which was performed in a Perkin Elmer DNA thermal cycler 480, consisted of denaturation at 94° C. for 1 min (in 1st cycle 1 min 30 sec), annealing at 60° C. for 1 min and elongation at 68° C. for 2 min. In one experiment a region from NS5B to the conserved region of the 3' UTR was amplified with the external primers H9261F and H3'X58R, and the internal primers H9282F and H3'X45R (Table 1). In another experiment, a segment of the variable region to the very end of the 3' UTR was amplified with the external primers H9375F and H3'X-35R, and the internal primers H9386F and H3'X-38R (Table 1, FIG. 1). Amplified products were purified with QIAquick PCR purification kit (from QIAGEN), digested with Hind III and Xba I (from Promega), purified by either gel electrophoresis, or phenol/chloroform extraction, and then cloned into the multiple cloning site of plasmid PGEM-9zf(-) (Promega) or pUC19 (Pharmacia). Cloning of cDNA into the vector was performed with T4 DNA ligase (Promega) by standard procedures.

Amplification of Near Full-Length H77 Genomes by Long PCR

The reactions were performed in a total volume of 50 μ l in 1 $\times$ buffer, 250 μ mol of each dNTP, 10 pmol each of sense and antisense primers, 1 μ l of the Advantage KlenTaq polymerase mix and 2 μ l of the cDNA reaction mixture (Tellier, R. et al. 1996 *PNAS USA* 93:4370-4373). A single PCR round of 35 cycles was performed in a Robocycler thermal cycler (from Stratagene), and consisted of denaturation at 99° C. for 35 sec, annealing at 67° C. for 30 sec and elongation at 68° C. for 10 min during the first 5 cycles, 11 min during the next 10 cycles, 12 min during the following 10 cycles and 13 min during the last 10 cycles. To amplify the complete ORF of HCV by long RT-PCR we used the sense primers H1 or A1 deduced from the 5' UTR and the antisense primer H9417R deduced from the variable region of the 3' UTR (Table 1, FIG. 1).

Construction of Full-Length H77 cDNA Clones

The long PCR products amplified with H1 and H9417R primers were cloned directly into pGEM-9zf(-) after digestion with Not I and Xba I (from Promega) (as per FIG. 1). Two clones were obtained with inserts of the expected size, pH21, and pH50.sub.I. Next, the chosen 3' UTR was cloned into both pH21, and pH50, after digestion with Afl II and Xba I (New England Biolabs). DH5 α competent cells (GIBCO/BRL) were transformed and selected with LB agar plates containing 100 μ g/ml ampicillin (from SIGMA). Then the selected colonies were cultured in LB liquid containing ampicillin at 30° C. for about 18-20 hrs (transformants containing full-length or near full-length cDNA of H77 produced a very low yield of plasmid when cultured at 37° C. or for more than 24 hrs). After small scale preparation (Wizard Plus Minipreps DNA Purification Systems, Promega) each plasmid was retransformed to select a single clone, and large scale preparation of plasmid DNA was performed with a QIAGEN plasmid Maxi kit.

Cloning of Long RT-PCR Products Into a Cassette Vector

To improve the efficiency of cloning, a vector with consensus 5' and 3' termini of HCV strain H77 was constructed (FIG. 1). This cassette vector (pCV) was obtained by cutting out the BamHI fragment (nts 1358-7530 of the H77 genome) from pH50, followed by religation. Next, the long PCR products of H77 amplified with H1 and H9417R or A1 and H9417R primers were purified (Geneclean spin kit; BIO 101) and cloned into pCV after digestion with AgeI and AflII (New England Biolabs) or with PinAI (isoschizomer of Age I) and BfrI (isoschizomer of Afl II) (Boehringer Mannheim). Large scale preparations of the plasmids containing full-length cDNA of H77 were performed as described above.

Construction of H77 Consensus Chimeric cDNA Clone

A full-length cDNA clone of H77 with an ORF encoding the consensus amino acid sequence was constructed by making a chimera from four of the cDNA clones obtained above. This consensus chimera, pCV-H77C, was constructed in two ligation steps by using standard molecular procedures and convenient cleavage sites and involved first a two piece ligation and then a three piece ligation. Large scale preparation of pCV-H77C was performed as already described.

In Vitro Transcription

Plasmids containing the full-length HCV cDNA were linearized with Xba I (from Promega), and purified by phenol/chloroform extraction and ethanol precipitation. A 100 μ l reaction mixture containing 10 μ g of linearized plasmid DNA, 1 $\times$ transcription buffer, 1 mM ATP, CTP, GTP and UTP, 10 mM DTT, 4% (v/v) RNasin (20-40 units/ μ l) and 2 μ l of T7 RNA polymerase (Promega) was incubated at 37° C. for 2 hrs. Five μ l of the reaction mixture was analyzed by agarose gel electrophoresis followed by ethidium bromide staining. The transcription reaction mixture was diluted with 400 μ l of ice-cold phosphate-buffered saline without calcium or magnesium, immediately frozen on dry ice and stored at -80° C. The final nucleic acid mixture was injected into chimpanzees within 24 hrs.

Intrahepatic Transfection of Chimpanzees

Laparotomy was performed and aliquots from two transcription reactions were injected into 6 sites of the exposed liver (Emerson, S. U. et al 1992 *J. Virol.* 66:6649-6654). Serum samples were collected weekly from chimpanzees and monitored for liver enzyme levels and anti-HCV antibodies. Weekly samples of 100 μ l of serum were tested for HCV RNA in a highly sensitive nested RT-PCR assay with AmpliTaq Gold (Perkin Elmer) (Yanagi, M. et al. 1996 *J. Infect. Dis.* 174:1324-1327; Bukh, J. et al. 1992 *PNAS USA* 89:187-191). The genome titer of HCV was estimated by testing 10-fold serial dilutions of the extracted RNA in the RT-PCR assay (Yanagi, M. et al. 1996 *J. Infect. Dis.* 174:1324-1327). The two chimpanzees used in this study were maintained under conditions that met all requirements for their use in an approved facility.

The consensus sequence of the complete ORF from HCV genomes recovered at week 2 post inoculation (p.i.) was determined by direct sequencing of PCR products obtained in long RT-PCR with primers A1 and H9417R followed by nested PCR of 10 overlapping fragments. The consensus sequence of the variable region of the 3' UTR was determined by direct sequencing of an amplicon obtained in nested RT-PCR as described above. Finally, we amplified selected regions independently by nested RT-PCR with AmpliTaq Gold.

Sequence Analysis

Both strands of DNA from PCR products, as well as plasmids, were sequenced with the ABI PRISM Dye Termination Cycle Sequencing Ready Reaction Kit using Taq DNA polymerase (Perkin Elmer) and about 100 specific sense and antisense sequence primers.

The consensus sequence of HCV strain H77 was determined in two different ways. In one approach, overlapping PCR products were directly sequenced, and amplified in nested RT-PCR from the H77 plasma sample. The sequence analyzed (nucleotides (nts) 35-9417) included the entire genome except the very 5' and 3' termini. In the second approach, the consensus sequence of nts 157-9384 was deduced from the sequences of 18 full-length cDNA clones.

Example 1

Variability in the Sequence of the 3' UTR of HCV Strain H77

The heterogeneity of the 3' UTR was analyzed by cloning and sequencing of DNA amplicons obtained in nested RT-PCR. 19 clones containing sequences of the entire variable region, the poly U-UC region and the adjacent 19 nt of the conserved region, and 65 clones containing sequences of the entire poly U-UC region and the first 63 nts of the conserved region were analyzed. This analysis confirmed that the variable region consisted of 43 nts, including two conserved termination codons (Han, J. M. et al 1992 *Nucl. Acids Res.* 20:3250). The sequence of the variable region was highly conserved within H77 since only 3 point mutations were found among the 19 clones analyzed. A poly U-UC region was present in all 84 clones analyzed. However, its length varied from 71-141 nts. The length of the poly U region was 9-103 nts, and that of the poly UC region was 35-85 nts. The number of C residues increased towards the 3' end of the poly UC region but the sequence of this region is not conserved. The first 63 nts of the conserved region were highly conserved among the clones analyzed, with a total of only 14 point mutations. To confirm the validity of the analysis, the 3' UTR was reamplified directly from a full-length cDNA clone of HCV (see below) by the nested-PCR procedure with the primers in the variable region and at the very 3' end of the HCV genome and cloned the PCR product. Eight clones had 1-7 nt deletions in the poly U region. Furthermore, although the C residues of the poly UC region were maintained, the spacing of these varied because of 1-2 nt deletions of U residues. These deletions must be artifacts introduced by PCR and such mistakes may have contributed to the heterogeneity originally observed in this region. However, the conserved region of the 3' UTR was amplified correctly, suggesting that the deletions were due to difficulties in transcribing a highly repetitive sequence.

One of the 3' UTR clones was selected for engineering of full-length cDNA clones of H77. This clone had the consensus variable sequence except for a single point mutation introduced to create an AflII cleavage site, a poly U-UC stretch of 81 nts with the most commonly observed UC pattern and the consensus sequence of the complete conserved region of 101 nts, including the distal 38 nts which originated from the antisense primer used in the amplification. After linearization with XbaI, the DNA template of this clone had the authentic 3' end.

Example 2

The Entire Open Reading Frame of H77 Amplified in One Round of Long RT-PCR

It had been previously demonstrated that a 9.25 kb fragment of the HCV genome from the 5' UTR to the 3' end of NS5B could be amplified from 104 GE (genome equivalents) of H77 by a single round of long RT-PCR (Tellier, R. et al. 1996 *J. Clin. Microbiol.* 34:3085-3091). In the current study, by optimizing primers and cycling conditions, the entire ORF of H77 was amplified in a single round of long RT-PCR with primers from the 5' UTR and the variable region of the 3' UTR. In fact, 9.4 kb of the H77 genome (H product: from the very 5' end to the variable region of the 3' UTR) could be amplified from 10⁵ GE or 9.3 kb (A product: from within the 5' UTR to the variable region of the 3' UTR) from 10⁴ GE or 10⁵ GE, in a single round of long RT-PCR (FIG. 2). The PCR products amplified from 10⁵ GE of H77 were used for engineering full-length cDNA clones (see below).

Example 3

Construction of Multiple Full-Length cDNA Clones of H77 in a Single Step by Cloning of Long RT-PCR Amplicons Directly into a Cassette Vector with Fixed 5' and 3' Termini

Direct cloning of the long PCR products (H), which contained a 5' T7 promoter, the authentic 5' end, the entire ORF of H77 and a short region of the 3' UTR, into pGEM-9zf(-) vector by NotI and XbaI digestion was first attempted. However, among the 70 clones examined all but two had inserts that were shorter than predicted. Sequence analysis identified a second Not I site in the majority of clones, which resulted in deletion of the nts past position 9221. Only two clones (pH21, and pH50) were missing the second NotI site and had the expected 5' and 3' sequences of the PCR product. Therefore, full-length cDNA clones (pH21 and pH50) were constructed by inserting the chosen 3' UTR into pH21₁ and pH50₁, respectively. Sequence analysis revealed that clone pH21 had a complete full-length sequence of H77; this clone was tested for infectivity. The second clone, pH50, had one nt deletion in the ORF at position 6365; this clone was used to make a cassette vector.

The complete ORF was amplified by constructing a cassette vector with fixed 5' and 3' termini as an intermediate of the full-length cDNA clones. This vector (pCV) was constructed by digestion of clone pH50 with BamHI, followed by religation, to give a shortened plasmid readily distinguished from plasmids containing the full-length insert. Attempts to clone long RT-PCR products (H) into pCV by AgeI and AflII yielded only 1 of 23 clones with an insert of the expected size. In order to increase the efficiency of cloning, we repeated the procedure but used PinAI and BfrI instead of the respective isoschizomers AgeI and AflII. By this protocol, 24 of 31 H clones and 30 of 35 A clones had the full-length cDNA of H77 as evaluated by restriction enzyme digestion. A total of 16 clones, selected at random, were each retransformed, and individual plasmids were purified and completely sequenced.

Demonstration of Infectious Nature of Transcripts of a cDNA Clone Representing the Consensus Sequence of Strain H77

A consensus chimera was constructed from 4 of the full-length cDNA clones with just 2 ligation steps. The final construct, pCV-H77C, had 11 nt differences from the consensus sequence of H77 in the ORF (FIG. 3). However, 10 of these nucleotide differences represented silent mutations. The chimeric clone differed from the consensus sequence at only one amino acid [L instead of F at position 790]. Among the 18 ORFs analyzed above, the F residue was found in 11 clones and the L residue in 7 clones. However, the L residue was dominant in other isolates of genotype 1a, including a first passage of H77 in a chimpanzee (Inchauspe, G. et al. 1991 *PNAS USA* 88:10292-10296).

To test the infectivity of the consensus chimeric clone of H77 intrahepatic transfection of a chimpanzee was performed. The pCV-H77C clone was linearized with Xba I and transcribed in vitro by T7 RNA polymerase (FIG. 2). The transcription mixture was next injected into 6 sites of the liver of chimpanzee 1530. The chimpanzee became infected with HCV as measured by detection of 10^2 GE/ml of viral genome at week 1 p.i. Furthermore, the HCV titer increased to 10^4 GE/ml at week 2 p.i., and reached 10^6 GE/ml by week 8 p.i. The viremic pattern observed in the early phase of the infection with the recombinant virus was similar to that observed in chimpanzees inoculated intravenously with strain H77 or other strains of HCV (Shimizu, Y. K. et al. 1990 *PNAS USA* 87:6441-6444).

The sequence of the HCV genomes from the serum sample collected at week 2 p.i. was analyzed. The consensus sequence of nts 298-9375 of the recovered genomes was determined by direct sequencing of PCR products obtained in long RT-PCR followed by nested PCR of 10 overlapping fragments. The identity to clone pCV-H77C sequence was 100%. The consensus sequence of nts 96-291, 1328-1848, 3585-4106, 4763-5113 and 9322-9445 was determined from PCR products obtained in different nested RT-PCR assays. The identity of these sequences with pCV-H77C was also 100%. These latter regions contained 4 mutations unique to the consensus chimera, including the artificial AflII cleavage site in the 3' UTR. Therefore, RNA transcripts of this clone of HCV were infectious.

The infectious nature of the consensus chimera indicates that the regions of the 5' and 3' UTRs incorporated into the cassette vector do not destroy viability. This makes it highly advantageous to use the cassette vector to construct infectious cDNA clones of other HCV strains when the consensus sequence for each ORF is inserted.

In addition, two complete full-length clones (dubbed pH21 and pCV-H11) constructed were not infectious, as shown by intrahepatic injection of chimpanzees with the corresponding RNA transcripts. Thus, injection of the transcription mixture into 3 sites of the exposed liver resulted in no observable HCV replication and weekly serum samples were negative for HCV RNA at weeks 1-17 p.i. in a highly sensitive nested RT-PCR assay. The cDNA template injected along with the RNA transcripts was also not detected in this assay.

Moreover, the chimpanzee remained negative for antibodies to HCV throughout the follow-up. Subsequent sequence analysis revealed that 7 of 16 additional clones were defective for polyprotein synthesis and all clones had multiple amino acid mutations compared with the consensus sequence of the parent strain. For example, clone pH21, which was not infec-

tious, had 7 amino acid substitutions in the entire predicted polyprotein compared with the consensus sequence of H77 (FIG. 3). The most notable mutation was at position 1026, which changed L to Q, altering the cleavage site between NS2 and NS3 (Reed, K. E. et al. 1995 *J. Virol.* 69:4127-4136). Clone pCV-H11, also non-infectious, had 21 amino acid substitutions in the predicted polyprotein compared with the consensus sequence of H77 (FIG. 3). The amino acid mutation at, position 564 eliminated a highly conserved C residue in the E2 protein (Okamoto, H. et al. 1992a *Virology* 188:331-341).

Example 4A

The chimpanzee of Example 4, designated 1530, was monitored out to 32 weeks p.i. for serum enzyme levels (ALT) and the presence of anti-HCV antibodies, HCV RNA, and liver histopathology. The results are shown in FIG. 18B.

A second chimp, designated 1494, was also transfected with RNA transcripts of the pCV-H77C clone and monitored out to 17 weeks p.i. for the presence of anti-HCV antibodies, HCV RNA and elevated serum enzyme levels. The results are shown in FIG. 18A.

Materials and Methods for Examples 5-10

Source of HCV Genotype 1b

An infectious plasma pool (second chimpanzee passage) containing strain HC-J4, genotype 1b, was prepared from acute phase plasma of a chimpanzee experimentally infected with serum containing HC-J4/91 (Okamoto, H. et al. 1992b *Virology* 190:894-899). The HC-J4/91 sample was obtained from a first chimpanzee passage during the chronic phase of hepatitis C about 8 years after experimental infection. The consensus sequence of the entire genome, except for the very 3' end, was determined previously for HC-J4/91 (Okamoto, H. et al. 1992b *Virology* 190:894-899).

Preparation of HCV RNA

Viral RNA was extracted from 100 μ l aliquots of the HC-J4 plasma pool with the TRIzol system (GIBCO BRL). The RNA pellets were each resuspended in 10 μ l of 10 mM dithiothreitol (DTT) with 5% (vol/vol) RNasin (20-40 units/ μ l) (Promega) and stored at -80° C. or immediately used for cDNA synthesis.

Amplification and Cloning of the 3' UTR

A region spanning from NS5B to the conserved region of the 3' UTR was amplified in nested RT-PCR using the procedure of Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743.

In brief, the RNA was denatured at 65° C. for 2 minutes, and cDNA was synthesized at 42° C. for 1 hour with Superscript II reverse transcriptase (GIBCO BRL) and primer H3'X58R (Table 1) in a 20 μ l reaction volume. The cDNA mixture was treated with RNase H and RNase T1 (GIBCO BRL) at 37° C. for 20 minutes. The first round of PCR was performed on 2 μ l of the final cDNA mixture in a total volume of 50 μ l with the Advantage cDNA polymerase mix (Clontech) and external primers H9261F (Table 1) and H3'X58R (Table 1). In the second round of PCR [internal primers H9282F (Table 1) and H3'X45R (Table 1)], 5 μ l of the first round PCR mixture was added to 45 μ l of the PCR reaction mixture. Each round of PCR (35 cycles), was performed in a DNA thermal cycler 480 (Perkin Elmer) and consisted of denaturation at 94° C. for 1 minute (1st cycle: 1 minute 30 sec), annealing at 60° C. for 1 minute and elongation at 68° C. for 2 minutes. After purification with QIAquick PCR purification kit (QIAGEN), digestion with HindIII and XbaI (Promega), and phenol/chloroform extraction, the amplified

products were cloned into pGEM-9zf(-) (Promega) (Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743).

Amplification and Cloning of the Entire ORF

A region from within the 5' UTR to the variable region of the 3' UTR of strain HC-J4 was amplified by long RT-PCR (FIG. 1) (Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743). The cDNA was synthesized at 42° C. for 1 hour in a 20 µl reaction volume with Superscript II reverse transcriptase and primer J4-9405R (5'-GCCTATTGGCCTGGAGTGGT-TAGCTC-3') (SEQ ID NO: 18), and treated with RNases as above. The cDNA mixture (2 µl) was amplified by long PCR with the Advantage cDNA polymerase mix and primers A1 (Table 1) (Bukh, J. et al. 1992 *PNAS USA* 89:187-191; Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743) and J4-9398R (5'-AGGATGGCCTTAAGGCCTGGAGTGGT-TAGCTCCCCGTTCA-3') (SEQ ID NO: 19). Primer J4-9398R contained extra bases (bold) and an artificial AflII cleavage site (underlined). A single PCR round was performed in a Robocycler thermal cycler (Stratagene), and consisted of denaturation at 99° C. for 35 seconds, annealing at 67° C. for 30 seconds and elongation at 68° C. for 10 minutes during the first S cycles, 11 minutes during the next 10 cycles, 12 minutes during the following 10 cycles and 13 minutes during the last 10 cycles.

After digesting the long PCR products obtained from strain HC-J4 with PinAI (isoschizomer of AgeI) and BfrI (isoschizomer of AflII) (Boehringer Mannheim), attempts were made to clone them directly into a cassette vector (pCV), which contained the 5' and 3' termini of strain H77 (FIG. 1) but no full-length clones were obtained. Accordingly, to improve the efficiency of cloning, the PCR product was further digested with BglII (Boehringer Mannheim) and the two resultant genome fragments [L fragment: PinAI/BglII, nts 156-8935; S fragment: BglII/BrfI, nts 8936-9398] were separately cloned into pCV (FIG. 6).

DH5.alpha. competent cells (GIBCO BRL) were transformed and selected on LB agar plates containing 100 µg/ml ampicillin (SIGMA) and amplified in LB liquid cultures at 30° C. for 18-20 hours.

Sequence analysis of 9 plasmids containing the S fragment (miniprep samples) and 9 plasmids containing the L fragment (maxiprep samples) were performed as described previously (Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743). Three L fragments, each encoding a distinct polypeptide, were cloned into pCV-J4S9 (which contained an S fragment encoding the consensus amino acid sequence of HC-J4) to construct three chimeric full-length HCV cDNAs (pCV-J4L2S, pCV-J4L4S and pCV-J4L6S) (FIG. 6). Large scale preparation of each clone was performed as described previously with a QIAGEN plasmid Maxi kit (Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743) and the authenticity of each clone was confirmed by sequence analysis.

Sequence Analysis

Both strands of DNA were sequenced with the ABI PRISM Dye Termination Cycle Sequencing Ready Reaction Kit using Taq DNA polymerase (Perkin Elmer) and about 90 specific sense and antisense primers. Analyses of genomic sequences, including multiple sequence alignments and tree analyses, were performed with GeneWorks (Oxford Molecular Group) (Bukh, J. et al. 1995 *Semin. Liver Dis.* 15:41-63).

The consensus sequence of strain HC-J4 was determined by direct sequencing of PCR products (nts 11-9412) and by sequence analysis of multiple cloned L and S fragments (nts 156-9371). The consensus sequence of the 3' UTR (3' variable

region, polypyrimidine tract and the first 16 nucleotides of the conserved region) was determined by analysis of 24 cDNA clones.

Intrahepatic Transfection of a Chimpanzee with Transcribed RNA

Two in vitro transcription reactions were performed with each of the three full-length clones. In each reaction 10 µg of plasmid DNA linearized with XbaI (Promega) was transcribed in a 100 µl reaction volume with T7 RNA polymerase (Promega) at 37° C. for 2 hours as described previously (Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743). Five µl of the final reaction mixture was analyzed by agarose gel electrophoresis and ethidium bromide staining (FIG. 5). Each transcription mixture was diluted with 400 µl of ice-cold phosphate-buffered saline without calcium or magnesium and then the two aliquots from the same cDNA clone were combined, immediately frozen on dry ice and stored at -80° C. Within 24 hours after freezing the transcription mixtures were injected into the chimpanzee by percutaneous intrahepatic injection that was guided by ultrasound. Each inoculum was individually injected (5-6 sites) into a separate area of the liver to prevent complementation or recombination. The chimpanzee was maintained under conditions that met all requirements for its use in an approved facility.

Serum samples were collected weekly from the chimpanzee and monitored for liver enzyme levels and anti-HCV antibodies. Weekly samples of 100 µl of serum were tested for HCV RNA in a sensitive nested RT-PCR assay (Bukh, J. et al. 1992 *PNAS USA* 89:187-191; Yanagi, M. et al. 1996 *J. Infect. Dis.* 174:1324-1327) with AmpliTaq Gold DNA polymerase. The genome equivalent (GE) titer of HCV was determined by testing 10-fold serial dilutions of the extracted RNA in the RT-PCR assay (Yanagi, M. et al. 1996 *J. Infect. Dis.* 174:1324-1327) with 1 GE defined as the number of HCV genomes present in the highest dilution which was positive in the RT-nested PCR assay.

To identify which of the three clones was infectious in vivo, the NS3 region (nts 3659-4110) from the chimpanzee serum was amplified in a highly sensitive and specific nested RT-PCR assay with AmpliTaq Gold DNA polymerase and the PCR products were cloned with a TA cloning kit (Invitrogen). In addition, the consensus sequence of the nearly complete genome (nts 11-9441) was determined by direct sequencing of overlapping PCR products.

Example 5

Sequence Analysis of Infectious Plasma Pool of Strain HC-J4 Used as the Cloning Source

As an infectious cDNA clone of a genotype 1a strain of HCV had been obtained only after the ORF was engineered to encode the consensus polypeptide (Kolykhalov, A. A., et al. 1997 *Science* 277:570-574; Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743), a detailed sequence analysis of the cloning source was performed to determine the consensus sequence prior to constructing an infectious cDNA clone of a 1b genotype.

A plasma pool of strain HC-J4 was prepared from acute phase plasmapheresis units collected from a chimpanzee experimentally infected with HC-J4/91 (Okamoto, H. et al. 1992 *Virology* 190:894-899). This HCV pool had a PCR titer of 10⁴-10⁵ GE/ml and an infectivity titer of approximately 10³ chimpanzee infectious doses per ml.

The heterogeneity of the 3' UTR of strain HC-J4 was determined by analyzing 24 clones of nested RT-PCR product. The

25

consensus sequence was identical to that previously published for HC-J4/91 (Okamoto, H. et al. 1992 *Virology* 190: 894-899), except at position 9407 (see below). The variable region consisted of 41 nucleotides (nts. 9372-9412), including two in-frame termination codons. Furthermore, its sequence was highly conserved except at positions 9399 (19 A and 5 T clones) and 9407 (17 T and 7 A clones). The poly U-UC region varied slightly in composition and greatly in length (31-162 nucleotides). In the conserved region, the first 16 nucleotides of 22 clones were identical to those previously published for other genotype 1 strains, whereas two clones each had a single point mutation. These data suggested that the structural organization at the 3' end of HC-J4 was similar to that of the infectious clone of a genotype 1a strain of Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743.

Next, the entire ORF of HC-J4 was amplified in a single round of long RT-PCR (FIG. 5). The original plan was to clone the resulting PCR products into the *PinAI* and *BrfI* site of a HCV cassette vector (pCV), which had fixed 5' and 3' termini of genotype 1 a (Yanagi et al., 1997) but since full-length clones were not obtained, two genome fragments (L and S) derived from the long RT-PCR products (FIG. 6) were separately subcloned into pCV.

To determine the consensus sequence of the ORF, the sequence of 9 clones each of the L fragment (pCV-J4L) and of the S fragment (pCV-J4S) was determined and quasispecies were found at 275 nucleotide (3.05%) and 78 amino acid (2.59%) positions, scattered throughout the 9030 nts (3010 aa) of the ORF (FIG. 7). Of the 161 nucleotide substitutions unique to a single clone, 71% were at the third position of the codon and 72% were silent.

Each of the nine L clones represented the near complete ORF of an individual genome. The differences among the L clones were 0.30-1.53% at the nucleotide and 0.31-1.47% at the amino acid level (FIG. 8). Two clones, L1 and L7, had a defective ORF due to a single nucleotide deletion and a single nucleotide insertion, respectively. Even though the HC-J4 plasma pool was obtained in the early acute phase, it appeared to contain at least three viral species (FIG. 9). Species A contained the L1, L2, L6, L8 and L9 clones, species B the L3, L7 and L10 clones and species C the L4 clone. Although each species A clone was unique all A clones differed from all B clones at the same 20 amino acid sites and at these positions, species C had the species A sequence at 14 positions and the species B sequence at 6 positions (FIG. 7).

Okamoto and coworkers (Okamoto, H. et al. 1992 *Virology* 190:894-899) previously determined the nearly complete genome consensus sequence of strain HC-J4 in acute phase serum of the first chimpanzee passage (HC-J4/83) as well as in chronic phase serum collected 8.2 years later (HC-J4/91). In addition, they determined the sequence of amino acids 379 to 413 (including HVR1) and amino acids 468 to 486 (including HVR2) of multiple individual clones (Okamoto, H. et al. 1992 *Virology* 190:894-899).

It was found by the present inventors that the sequences of individual genomes in the plasma pool collected from a chimpanzee inoculated with HC-J4/91 were all more closely related to HC-J4/91 than to HC-J4/83 (FIGS. 8, 9) and contained HVR amino acid sequences closely related to three of the four viral species previously found in HC-J4/91 (FIG. 10).

Thus, the data presented herein demonstrate the occurrence of the simultaneous transmission of multiple species to a single chimpanzee and clearly illustrates the difficulties in accurately determining the evolution of HCV over time since multiple species with significant changes throughout the HCV genome can be present from the onset of the infection. Accordingly, infection of chimpanzees with monoclonal

26

viruses derived from the infectious clones described herein will make it possible to perform more detailed studies of the evolution of HCV in vivo and its importance for viral persistence and pathogenesis.

Example 6

Determination of the Consensus Sequence of HC-J4 in the Plasma Pool

The consensus sequence of nucleotides 156-9371 of HC-J4 was determined by two approaches. In one approach, the consensus sequence was deduced from 9 clones of the long RT-PCR product. In the other approach the long RT-PCR product was reamplified by PtR as overlapping fragments which were sequenced directly. The two "consensus" sequences differed at 31 (0.34%) of 9216 nucleotide positions and at 11 (0.37%) of 3010 deduced amino acid positions (FIG. 7). At all of these positions a major quasispecies of strain HC-J4 was found in the plasma pool. At 9 additional amino acid positions the cloned sequences displayed heterogeneity and the direct sequence was ambiguous (FIG. 7). Finally, it should be noted that there were multiple amino acid positions at which the consensus sequence obtained by direct sequencing was identical to that obtained by cloning and sequencing even though a major quasispecies was detected (FIG. 7).

For positions at which the two "consensus" sequences of HC-J4 differed, both amino acids were included in a composite consensus sequence (FIG. 7). However, even with this allowance, none of the 9 L clones analyzed (aa 1-2864) had the composite consensus sequence: two clones did not encode the complete polypeptide and the remaining 7 clones differed from the consensus sequence by 3-13 amino acids (FIG. 7).

Example 7

Construction of Chimeric Full-Length cDNA Clones Containing The Entire ORF of HC-J4

The cassette vector used to clone strain H77 was used to construct an infectious cDNA clone containing the ORF of a second subtype.

In brief, three full-length cDNA clones were constructed by cloning different L fragments into the *PinAI*/*BglIII* site of pCV-J4S9, the cassette vector for genotype 1 a (FIG. 6), which also contained an S fragment encoding the consensus amino acid sequence of HC-J4. Therefore, although the ORF was from strain HC-J4, most of the 5' and 3' terminal sequences originated from strain H77. As a result, the 5' and 3' UTR were chimeras of genotypes 1a and 1b (FIG. 11).

The first 155 nucleotides of the 5' UTR were from strain H77 (genotype 1a), and differed from the authentic sequence of HC-J4 (genotype 1) at nucleotides 11, 12, 13, 34 and 35. In two clones (pCV-J4L2S, pCV-J4L6S) the rest of the 5' UTR had the consensus sequence of HC-J4, whereas the third clone (pCV-J4L4S) had a single nucleotide insertion at position 207. In all 3 clones the first 27 nucleotides of the 3' variable region of the 3' UTR were identical with the consensus sequence of HC-J4. The remaining 15 nucleotides of the variable region, the poly U-UC region and the 3' conserved region of the 3' UTR had the same sequence as an infectious clone of strain H77 (FIG. 11).

None of the three full-length clones of HC-J4 had the ORF composite consensus sequence (FIGS. 7, 12). The pCV-J4L6S clone had only three amino acid changes: Q for R at position 231 (E1), V for A at position 937 (NS2) and T for S

at position 1215 (NS3). The pCV-J4L4S clone had 7 amino acid changes, including a change at position 450 (E2) that eliminated a highly conserved N-linked glycosylation site (Okamoto, H. et al. 1992 *Virology* 188:331-341). Finally, the pCV-J4L2S clone had 9 amino acid changes compared with the consensus sequence of HC-J4. A change at position 304 (E1) mutated a highly conserved cysteine residue (Bukh, J. et al. 1993 *PNAS USA* 90:8234-8238; Okamoto, H. et al. 1992 *Virology* 188:331-341).

Example 8

Transfection of a Chimpanzee by In Vitro Transcripts of a Chimeric cDNA

The infectivity of the three chimeric HCV clones was determined by ultra-sound-guided percutaneous intrahepatic injection into the liver of a chimpanzee of the same amount: of cDNA and transcription mixture for each of the clones (FIG. 5). This procedure is a less invasive procedure than the laparotomy procedure utilized by Kolykhalov, A. A., et al. 1997 *Science* 277:570-574 and Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743 and should facilitate in vivo studies of cDNA clones of HCV in chimpanzees since percutaneous procedures, unlike laparotomy, can be performed repeatedly.

As shown in FIG. 13, the chimpanzee became infected with HCV as measured by increasing titers of 10^2 GE/ml at week 1 p.i., 10^3 GE/ml at week 2 p.i. and 10^4 - 10^5 GE/ml during weeks 3 to 10 p.i.

The viremic pattern found in the early phase of the infection was similar to that observed for the recombinant H77 virus in chimpanzees (Bukh et al., unpublished data; Kolykhalov, A. A., et al. 1997 *Science* 277:570-574; Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743). The chimpanzee transfected in the present study was chronically infected with hepatitis G virus (HGV/GBV-C) (Bukh, J. et al. 1998 Experimental infection of chimpanzees with hepatitis G virus and genetic analysis of the virus. *J. Infect. Dis.* (in press) and had a titer of 10^6 GE/ml at the time of HCV transfection. Although HGV/GBV-C was originally believed to be a hepatitis virus, it does not cause hepatitis in chimpanzees (Bukh, J. et al. 1998 Experimental infection of chimpanzees with hepatitis G virus and genetic analysis of the virus. *J. Infect. Dis.* (in press) and may not replicate in the liver (Laskus, T. et al. 1997 *J. Virol.* 71:7804-7806). The present study demonstrated that an ongoing infection of HGV/GBV-C did not prevent acute HCV infection in the chimpanzee model.

However, to identify which of the three full-length HC-J4 clones were infectious, the NS3 region (nts. 3659-4110) of HCV genomes amplified by RT-PCR from serum samples taken from the infected chimpanzee during weeks 2 and 4 post-infection (p.i.) were cloned and sequenced. As the PCR primers were a complete match with each of the original three clones, this assay should not have preferentially amplified one virus over another. Sequence analysis of 26 and 24 clones obtained at weeks 2 and 4 p.i., respectively, demonstrated that all originated from the transcripts of pCV-J4L6S.

Moreover, the consensus sequence of PCR products of the nearly complete genome (nts. 11-9441), amplified from serum obtained during week 2 p.i., was identical to the sequence of pCV-J4L6S and there was no evidence of quasispecies. Thus, RNA transcripts of pCV-J4L6S, but not of pCV-J4L2S or pCV-J4L4S, were infectious in vivo. The data in FIG. 13 is therefore the product of the transfection of RNA transcripts of pCV-J4L6S.

In addition, the chimeric sequences of genotypes 1a and 1b in the UTRs were maintained in the infected chimpanzee. The

consensus sequence of nucleotides 11-341 of the 5' UTR and the variable region of the 3' UTR, amplified from serum obtained during weeks 2 and 4 p.i., had the expected chimeric sequence of genotypes 1a and 1b (FIG. 11). Also three of four clones of the 3' UTR obtained at week 2 p.i. had the chimeric sequence of the variable region, whereas a single substitution was noted in the fourth clone. However, in all four clones the poly U region was longer (2-12 nts) than expected. Also, extra C and G residues were observed in this region. For the most part, the number of C residues in the poly UC region was maintained in all clones although the spacing varied. As shown previously, variations in the number of U residues can reflect artifacts introduced during PCR amplification (Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743). The sequence of the first 19 nucleotides of the conserved region was maintained in all four clones. Thus, with the exception of the poly U-UC region, the genomic sequences recovered from the infected chimpanzee were exactly those of the chimeric infectious clone pCV-J4BL6S.

The results presented in FIG. 13 therefore demonstrate that HCV polypeptide sequences other than the consensus sequence can be infectious and that a chimeric genome containing portions of the H77 termini could produce an infectious virus. In addition, these results showed for the first time that it is possible to make infectious viruses containing 5' and 3' terminal sequences specific for two different subtypes of the same major genotype of HCV.

Example 9

Construction of a Chimeric 1a/1b Infectious Clone

A chimeric 1a/1b infectious clone in which the structural region of the genotype 1b infectious clone is inserted into the 1a clone of Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743 is constructed by following the protocol shown in FIG. 15. The resultant chimera contains nucleotides 156-2763 of the 1b clone described herein inserted into the 1a clone of FIGS. 4A-4F. The sequences of the primers shown in FIG. 15 which are used in constructing this chimeric clone, designated pH77CV-J4, are presented below.

1. H2751S (ClaI/Nde I) (SEQ ID NO: 20)
CGT CAT CGA TCC TCA GCG GGC ATA TGC ACT GGA
CAC GGA
2. H2870R (SEQ ID NO: 21)
CAT GCA CCA GCT GAT ATA GCG CTT GTA ATA TG
3. H7851S (SEQ ID NO: 22)
TCC GTA GAG GAA GCT TGC AGC CTG ACG CCC
4. H9173 R (P-M) (SEQ ID NO: 23)
GTA CTT GCC ACA TAT AGC AGC CCT GCC TCC TCT G
5. H9140S (P-M) (SEQ ID NO: 24)
CAG AGG AGG CAG GGC TGC TAT ATG TGG CAA GTA C
6. H9417R (SEQ ID NO: 25)
CGT CTC TAG ACA GGA AAT GGC TTA AGA GGC CGG
AGT GTT TAC C
7. J4-2271S (SEQ ID NO: 26)
TGC AAT TGG ACT CGA GGA GAG CGC TGT AAC TTG
GAG
8. J4-2776R (Nde I) (SEQ ID NO: 27)
CGG TCC AAG GCA TAT GCT CGT GGT GGT AAC GCC
AG

Transcripts of the chimeric 1a/1b clone (whose sequence is shown in FIGS. 16A-16F) are then produced and transfected

into chimpanzees by the methods described in the Materials and Methods section herein and the transfected animals are then be subjected to biochemical (ALT levels), histopathological and PCR analyses to determine the infectivity of the chimeric clone.

Example 10

Construction of 3' Deletion Mutants of the 1a Infectious Clone pCV-H77C

Seven constructs having various deletions in the 3' untranslated region (UTR) of the 1a infectious clone pCV-H77C were constructed as described in FIGS. 17A-17B. The 3' untranslated sequence remaining in each of the seven constructs following their respective deletions is shown in FIGS. 17A-17B.

Construct pCV-H77C(-98X) containing a deletion of the 3'-most 98 nucleotide sequences in the 3'-UTR was transcribed in vitro according to the methods described herein and 1 ml of the diluted transcription mixture was percutaneously transfected into the liver of a chimpanzee with the aid of ultrasound. After three weeks, the transfection was repeated. The chimpanzee was observed to be negative for hepatitis C virus replication as measured by RT-PCR assay for 5 weeks after transfection. These results demonstrate that the deleted 98 nucleotide 3'-UTR sequence was critical for production of infectious HCV and appear to contradict the reports of Dash, S. et al. 1997 *Am. J. Pathol.* 151: 363-373 and Yoo, B. J. et al. 1995 *J. Virol.* 69:32-38 who reported that RNA transcripts from cDNA clones of HCV-1 and HCV-N lacking the terminal 98 conserved nucleotides at the very 3' end of the 3'-UTR resulted in viral replication after transfection into human hematoma cell lines.

Transcripts of the (-42X) mutant (FIG. 17C) were also produced and transfected into a chimpanzee and transcripts of the other five deletion mutants shown in FIGS. 17D-17G) are to be produced and transfected into chimpanzees by the methods described herein. All transfected animals are to then be assayed for viral replication via RT-PCR.

Discussion

In two recent reports on transfection of chimpanzees, only those clones engineered to have the independently determined and slightly different consensus amino acid sequence of the polypeptide of strain H77 were infectious (Kolykhalov, A. A., et al. 1997 *Science* 277:570-574; Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743). Although the two infectious clones differed at four amino acid positions, these differences were represented in a major component of the quasispecies of the cloning source. In the present study, a single consensus sequence of strain HC-J4 could not be defined because the consensus sequence obtained by two different approaches (direct sequencing and sequencing of cloned products) differed at 20 amino acid positions, even though the same genomic PCR product was analyzed. The infectious clone differed at two positions from the composite amino acid consensus sequence, from the sequence of the 8 additional HC-J4 clones analyzed in this study and from published sequences of earlier passage samples. An additional amino acid differed from the composite consensus sequence but was found in two other HC-J4 clones analyzed in this study. The two non-infectious full-length clones of HC-J4 differed from the composite consensus sequence by only 7 and 9 amino acid differences. However, since these clones had the same termini as the infectious clone (except for a single nucleotide inser-

tion in the 5' UTR of pCV-J4L4S), one or more of these amino acid changes in each clone was apparently deleterious for the virus.

It was also found in the present study that HC-J4, like other strains of genotype 1b (Kolykhalov, A. A. et al. 1996 *J. Virol.* 70:3363-3371; Tanaka, T. et al. 1996 *J. Virol.* 70:3307-3312; Yamada, N. et al. 1996 *Virology* 223:255-261), had a poly U-UC region followed by a terminal conserved element. The poly U-UC region appears to vary considerably so it was not clear whether changes in this region would have a significant effect on virus replication. On the other hand, the 3' 98 nucleotides of the HCV genome were previously shown to be identical among other strains of genotypes 1a and 1b (Kolykhalov, A. A. et al. 1996 *J. Virol.* 70:3363-3371; Tanaka, T. et al. 1996 *J. Virol.* 70:3307-3312). Thus, use of the cassette vector would not alter this region except for addition of 3 nucleotides found in strain H77 between the poly UC region and the 3' 98 conserved nucleotides.

In conclusion, an infectious clone representing a genotype 1b strain of HCV has been constructed. Thus, it has been demonstrated that it was possible to obtain an infectious clone of a second strain of HCV. In addition, it has been shown that a consensus amino acid sequence was not absolutely required for infectivity and that chimeras between the UTRs of two different genotypes could be viable.

REFERENCES

- Alter, M. J. 1997 *Hepatology* 26:62S-65S.
- Blight, K. J. and Rice, C. M. 1997 *J. Virol.* 71:7345-7352.
- Brechot, C. 1997 *Hepatology* 25:772-774.
- Bukh, J. et al. 1992 *PNAS USA* 89:187-191.
- Bukh, J. et al. 1993 *PNAS USA* 90:8234-8238.
- Bukh, J. et al. 1995 *Semin. Liver Dis.* 15:41-63.
- Bukh, J. et al. 1997 Genetic heterogeneity of hepatitis C virus and related viruses. In "Viral Hepatitis and Liver Disease, Proceedings of IX Triennial International Symposium on Viral Hepatitis and Liver Disease, Rome, Italy, 1996" (M. Rizzetto, R. H. Purcell, J. L. Gerin and G. Verme, Eds.), pp. 167-175. Edizioni Minerva Medica, Turin.
- Bukh, J. et al. 1998 Experimental infection of chimpanzees with hepatitis G virus and genetic analysis of the virus. *J. Infect. Dis.* (in press).
- Choo, Q.-L. et al. 1991 *PNAS USA* 88:2451-2455.
- Dash, S. et al. 1997 *Am. J. Pathol.* 151: 363-373.
- Emerson, S. U. et al 1992 *J. Virol.* 66:6649-6654.
- Farci, P. and Purcell, R. H. 1993 Natural history and experimental models. In "Viral Hepatitis: Scientific Basis and Clinical Management". (A. J. Zuckerman and H. C. Thomas, Eds.), pp. 241-267. Churchill Livingstone, Edinburgh.
- Farci, P. et al. 1994 Prevention of hepatitis C virus infection in chimpanzees after antibody-mediated in vitro neutralization. *PNAS USA* 91:7792-7796.
- Farci, P. et al. 1996 *PNAS USA* 93:15394-15399.
- Farci, P. et al. 1997 *Springer Semin. Immunopathol.* 19:5-26.
- Fausto, N. 1997 *Am. J. Pathol.* 151:361.
- Feinstone, S. M. et al 1981 *J. Infect. Dis.* 144:588-598.
- Fried, M. W. and Hoofnagle, J. H. 1995 *Semin. Liver Dis.* 15:82-91.
- Han, J. M. et al 1992 *Nucl. Acids Res.* 20:3250.
- Hijikata, M. et al. 1991 *Biochem. Biophys. Res. Commun.* 175:220-228.
- Honda, M. et al. 1996 *RNA* 2:955-968.
- Hoofnagle, J. H. 1997 *Hepatology* 26:15S-20S.

31

Houghton, M. 1996 Hepatitis C viruses. In "Fields Virology" (B. N. Fields, D. M. Knipe, P. M. Howley, et al., Eds.), Third ed., Lippincott-Raven Publishers, Philadelphia.

Inchauspe, G. et al. 1991 *PNAS USA* 88:10292-10296.

Ito, T. and Lai, M. M. C. 1997 *J. Virol.* 71:8698-8706.

Kato, N., et al. 1990 *PNAS USA* 87:9524-9528.

Kolykhalov, A. A. et al. 1996 *J. Virol.* 70:3363-3371.

Kolykhalov, A. A., et al. 1997 *Science* 277:570-574.

Krieg, A. M. et al. 1995 *Nature* 374:546.

Krieg, A. M. et al. 1996 *J. Lab. Clin. Med.* 128:128.

Laskus, T. et al. 1997 *J. Virol.* 71:7804-7806.

Major, M. E. and Feinstone, S. M. 1997 *Heptology* 25:1527-1538.

Ogata, N. et al. 1991 *PNAS USA* 88:3392-3396.

Okamoto, H. et al. 1992a *Virology* 188:331-341.

Okamoto, H. et al. 1992b *Virology* 190:894-899.

Reed, K. E. et al. 1995 *J. Virol.* 69:4127-4136.

Rice, C. M. 1996 Flaviviridae: The viruses and their replication, In "Fields Virology". (B. N. Fields, D. M. Knipe, P. M. Howley, et al., Eds.), Third ed., Lippincott-Raven Publishers, Philadelphia.

32

Shimizu, Y. K. et al. 1990 *PNAS USA* 87:6441-6444.

Shimizu, Y. K. et al. 1992 *PNAS USA* 89:5477-5481.

Shimizu, Y. K. et al. 1996 *Virology* 223:409-412.

Simmonds, P. et al. 1993 *J. Gen. Virol.* 74:2391-2399.

Takamizawa, A. et al. 1991 *J. Virol.* 65:1105-1113.

Tanaka, T. et al. 1995 *Biochem. Biophys. Res. Commun.* 215:744-749.

Tanaka, T. et al. 1996 *J. Virol.* 70:3307-3312.

Tellier, R. et al. 1996 *PNAS USA* 93:4370-4373.

Tellier, R. et al. 1996a *J. Clin. Microbiol.* 34:3085-3091.

Tsuchihara, K. et al. 1997 *J. Virol.* 71:6720-6726.

Tsukiyama-Kohara, K. et al. 1992 *J. Virol.* 66:1476-1483.

Weiner, A. J. et al. 1991 *Virology* 180:842-848.

Weiner, A. J. et al. 1992 *PNAS USA* 89:3468-3472.

Yamada, N. et al. 1996 *Virology* 223:255-261.

Yanagi, M. et al. 1996 *J. Infect. Dis.* 174:1324-1327.

Yanagi, M. et al. 1997 *PNAS USA* 94:8738-8743.

Yoo, B. J. et al. 1995 *J. Virol.* 69:32-38.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 65

<210> SEQ ID NO 1

<211> LENGTH: 3011

<212> TYPE: PRT

<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 1

```
Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn
 1           5           10           15

Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly
 20           25           30

Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala
 35           40           45

Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro
 50           55           60

Ile Pro Lys Ala Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly
 65           70           75           80

Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp
 85           90           95

Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro
100           105           110

Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys
115           120           125

Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu
130           135           140

Gly Gly Ala Ala Arg Ala Leu Ala His Gly Val Arg Val Leu Glu Asp
145           150           155           160

Gly Val Asn Tyr Ala Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile
165           170           175

Phe Leu Leu Ala Leu Leu Ser Cys Leu Thr Val Pro Ala Ser Ala Tyr
180           185           190

Gln Val Arg Asn Ser Ser Gly Leu Tyr His Val Thr Asn Asp Cys Pro
195           200           205

Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Ala Ile Leu His Thr Pro
```

-continued

210	215	220
Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ala Ser Arg Cys Trp Val 225 230 235 240		
Ala Val Thr Pro Thr Val Ala Thr Arg Asp Gly Lys Leu Pro Thr Thr 245 250 255		
Gln Leu Arg Arg His Ile Asp Leu Leu Val Gly Ser Ala Thr Leu Cys 260 265 270		
Ser Ala Leu Tyr Val Gly Asp Leu Cys Gly Ser Val Phe Leu Val Gly 275 280 285		
Gln Leu Phe Thr Phe Ser Pro Arg Arg His Trp Thr Thr Gln Asp Cys 290 295 300		
Asn Cys Ser Ile Tyr Pro Gly His Ile Thr Gly His Arg Met Ala Trp 305 310 315 320		
Asp Met Met Met Asn Trp Ser Pro Thr Ala Ala Leu Val Val Ala Gln 325 330 335		
Leu Leu Arg Ile Pro Gln Ala Ile Met Asp Met Ile Ala Gly Ala His 340 345 350		
Trp Gly Val Leu Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp 355 360 365		
Ala Lys Val Leu Val Val Leu Leu Leu Phe Ala Gly Val Asp Ala Glu 370 375 380		
Thr His Val Thr Gly Gly Asn Ala Gly Arg Thr Thr Ala Gly Leu Val 385 390 395 400		
Gly Leu Leu Thr Pro Gly Ala Lys Gln Asn Ile Gln Leu Ile Asn Thr 405 410 415		
Asn Gly Ser Trp His Ile Asn Ser Thr Ala Leu Asn Cys Asn Glu Ser 420 425 430		
Leu Asn Thr Gly Trp Leu Ala Gly Leu Phe Tyr Gln His Lys Phe Asn 435 440 445		
Ser Ser Gly Cys Pro Glu Arg Leu Ala Ser Cys Arg Arg Leu Thr Asp 450 455 460		
Phe Ala Gln Gly Trp Gly Pro Ile Ser Tyr Ala Asn Gly Ser Gly Leu 465 470 475 480		
Asp Glu Arg Pro Tyr Cys Trp His Tyr Pro Pro Arg Pro Cys Gly Ile 485 490 495		
Val Pro Ala Lys Ser Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser 500 505 510		
Pro Val Val Val Gly Thr Thr Asp Arg Ser Gly Ala Pro Thr Tyr Ser 515 520 525		
Trp Gly Ala Asn Asp Thr Asp Val Phe Val Leu Asn Asn Thr Arg Pro 530 535 540		
Pro Leu Gly Asn Trp Phe Gly Cys Thr Trp Met Asn Ser Thr Gly Phe 545 550 555 560		
Thr Lys Val Cys Gly Ala Pro Pro Cys Val Ile Gly Gly Val Gly Asn 565 570 575		
Asn Thr Leu Leu Cys Pro Thr Asp Cys Phe Arg Lys His Pro Glu Ala 580 585 590		
Thr Tyr Ser Arg Cys Gly Ser Gly Pro Trp Ile Thr Pro Arg Cys Met 595 600 605		
Val Asp Tyr Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Ile Asn Tyr 610 615 620		
Thr Ile Phe Lys Val Arg Met Tyr Val Gly Gly Val Glu His Arg Leu 625 630 635 640		

-continued

Glu	Ala	Ala	Cys	Asn	Trp	Thr	Arg	Gly	Glu	Arg	Cys	Asp	Leu	Glu	Asp
			645						650					655	
Arg	Asp	Arg	Ser	Glu	Leu	Ser	Pro	Leu	Leu	Leu	Ser	Thr	Thr	Gln	Trp
			660					665						670	
Gln	Val	Leu	Pro	Cys	Ser	Phe	Thr	Thr	Leu	Pro	Ala	Leu	Ser	Thr	Gly
		675					680							685	
Leu	Ile	His	Leu	His	Gln	Asn	Ile	Val	Asp	Val	Gln	Tyr	Leu	Tyr	Gly
	690					695					700				
Val	Gly	Ser	Ser	Ile	Ala	Ser	Trp	Ala	Ile	Lys	Trp	Glu	Tyr	Val	Val
705					710					715					720
Leu	Leu	Phe	Leu	Leu	Leu	Ala	Asp	Ala	Arg	Val	Cys	Ser	Cys	Leu	Trp
			725						730						735
Met	Met	Leu	Leu	Ile	Ser	Gln	Ala	Glu	Ala	Ala	Leu	Glu	Asn	Leu	Val
		740						745						750	
Ile	Leu	Asn	Ala	Ala	Ser	Leu	Ala	Gly	Thr	His	Gly	Leu	Val	Ser	Phe
	755						760						765		
Leu	Val	Phe	Phe	Cys	Phe	Ala	Trp	Tyr	Leu	Lys	Gly	Arg	Trp	Val	Pro
	770						775					780			
Gly	Ala	Val	Tyr	Ala	Leu	Tyr	Gly	Met	Trp	Pro	Leu	Leu	Leu	Leu	Leu
785					790					795					800
Leu	Ala	Leu	Pro	Gln	Arg	Ala	Tyr	Ala	Leu	Asp	Thr	Glu	Val	Ala	Ala
				805					810						815
Ser	Cys	Gly	Gly	Val	Val	Leu	Val	Gly	Leu	Met	Ala	Leu	Thr	Leu	Ser
			820					825						830	
Pro	Tyr	Tyr	Lys	Arg	Tyr	Ile	Ser	Trp	Cys	Met	Trp	Trp	Leu	Gln	Tyr
		835					840						845		
Phe	Leu	Thr	Arg	Val	Glu	Ala	Gln	Leu	His	Val	Trp	Val	Pro	Pro	Leu
	850						855					860			
Asn	Val	Arg	Gly	Gly	Arg	Asp	Ala	Val	Ile	Leu	Leu	Met	Cys	Val	Val
865					870					875					880
His	Pro	Thr	Leu	Val	Phe	Asp	Ile	Thr	Lys	Leu	Leu	Leu	Ala	Ile	Phe
			885						890					895	
Gly	Pro	Leu	Trp	Ile	Leu	Gln	Ala	Ser	Leu	Leu	Lys	Val	Pro	Tyr	Phe
		900						905						910	
Val	Arg	Val	Gln	Gly	Leu	Leu	Arg	Ile	Cys	Ala	Leu	Ala	Arg	Lys	Ile
		915					920						925		
Ala	Gly	Gly	His	Tyr	Val	Gln	Met	Ala	Ile	Ile	Lys	Leu	Gly	Ala	Leu
	930					935						940			
Thr	Gly	Thr	Tyr	Val	Tyr	Asn	His	Leu	Thr	Pro	Leu	Arg	Asp	Trp	Ala
945					950					955					960
His	Asn	Gly	Leu	Arg	Asp	Leu	Ala	Val	Ala	Val	Glu	Pro	Val	Val	Phe
			965						970						975
Ser	Arg	Met	Glu	Thr	Lys	Leu	Ile	Thr	Trp	Gly	Ala	Asp	Thr	Ala	Ala
			980					985						990	
Cys	Gly	Asp	Ile	Ile	Asn	Gly	Leu	Pro	Val	Ser	Ala	Arg	Arg	Gly	Gln
		995					1000					1005			
Glu	Ile	Leu	Leu	Gly	Pro	Ala	Asp	Gly	Met	Val	Ser	Lys	Gly	Trp	Arg
	1010					1015						1020			
Leu	Leu	Ala	Pro	Ile	Thr	Ala	Tyr	Ala	Gln	Gln	Thr	Arg	Gly	Leu	Leu
1025					1030					1035					1040
Gly	Cys	Ile	Ile	Thr	Ser	Leu	Thr	Gly	Arg	Asp	Lys	Asn	Gln	Val	Glu
			1045					1050						1055	

Gly	Glu	Val	Gln	Ile	Val	Ser	Thr	Ala	Thr	Gln	Thr	Phe	Leu	Ala	Thr
			1060						1065			1070			
Cys	Ile	Asn	Gly	Val	Cys	Trp	Thr	Val	Tyr	His	Gly	Ala	Gly	Thr	Arg
			1075						1080			1085			
Thr	Ile	Ala	Ser	Pro	Lys	Gly	Pro	Val	Ile	Gln	Met	Tyr	Thr	Asn	Val
			1090						1095			1100			
Asp	Gln	Asp	Leu	Val	Gly	Trp	Pro	Ala	Pro	Gln	Gly	Ser	Arg	Ser	Leu
			1105						1110			1115			
Thr	Pro	Cys	Thr	Cys	Gly	Ser	Ser	Asp	Leu	Tyr	Leu	Val	Thr	Arg	His
			1120						1125			1130			
Ala	Asp	Val	Ile	Pro	Val	Arg	Arg	Arg	Gly	Asp	Ser	Arg	Gly	Ser	Leu
			1140						1145			1150			
Leu	Ser	Pro	Arg	Pro	Ile	Ser	Tyr	Leu	Lys	Gly	Ser	Ser	Gly	Gly	Pro
			1155						1160			1165			
Leu	Leu	Cys	Pro	Ala	Gly	His	Ala	Val	Gly	Leu	Phe	Arg	Ala	Ala	Val
			1170						1175			1180			
Cys	Thr	Arg	Gly	Val	Ala	Lys	Ala	Val	Asp	Phe	Ile	Pro	Val	Glu	Asn
			1185						1190			1195			
Leu	Gly	Thr	Thr	Met	Arg	Ser	Pro	Val	Phe	Thr	Asp	Asn	Ser	Ser	Pro
			1200						1205			1210			
Pro	Ala	Val	Pro	Gln	Ser	Phe	Gln	Val	Ala	His	Leu	His	Ala	Pro	Thr
			1215						1220			1225			
Gly	Ser	Gly	Lys	Ser	Thr	Lys	Val	Pro	Ala	Ala	Tyr	Ala	Ala	Gln	Gly
			1230						1235			1240			
Tyr	Lys	Val	Leu	Val	Leu	Asn	Pro	Ser	Val	Ala	Ala	Thr	Leu	Gly	Phe
			1245						1250			1255			
Gly	Ala	Tyr	Met	Ser	Lys	Ala	His	Gly	Val	Asp	Pro	Asn	Ile	Arg	Thr
			1260						1265			1270			
Gly	Val	Arg	Thr	Ile	Thr	Thr	Gly	Ser	Pro	Ile	Thr	Tyr	Ser	Thr	Tyr
			1275						1280			1285			
Gly	Lys	Phe	Leu	Ala	Asp	Gly	Gly	Cys	Ser	Gly	Gly	Ala	Tyr	Asp	Ile
			1290						1295			1300			
Ile	Ile	Cys	Asp	Glu	Cys	His	Ser	Thr	Asp	Ala	Thr	Ser	Ile	Leu	Gly
			1305						1310			1315			
Ile	Gly	Thr	Val	Leu	Asp	Gln	Ala	Glu	Thr	Ala	Gly	Ala	Arg	Leu	Val
			1320						1325			1330			
Val	Leu	Ala	Thr	Ala	Thr	Pro	Pro	Gly	Ser	Val	Thr	Val	Ser	His	Pro
			1335						1340			1345			
Asn	Ile	Glu	Glu	Val	Ala	Leu	Ser	Thr	Thr	Gly	Glu	Ile	Pro	Phe	Tyr
			1350						1355			1360			
Gly	Lys	Ala	Ile	Pro	Leu	Glu	Val	Ile	Lys	Gly	Gly	Arg	His	Leu	Ile
			1365						1370			1375			
Phe	Cys	His	Ser	Lys	Lys	Lys	Cys	Asp	Glu	Leu	Ala	Ala	Lys	Leu	Val
			1380						1385			1390			
Ala	Leu	Gly	Ile	Asn	Ala	Val	Ala	Tyr	Tyr	Arg	Gly	Leu	Asp	Val	Ser
			1395						1400			1405			
Val	Ile	Pro	Thr	Ser	Gly	Asp	Val	Val	Val	Val	Ser	Thr	Asp	Ala	Leu
			1410						1415			1420			
Met	Thr	Gly	Phe	Thr	Gly	Asp	Phe	Asp	Ser	Val	Ile	Asp	Cys	Asn	Thr
			1425						1430			1435			
Cys	Val	Thr	Gln	Thr	Val	Asp	Phe	Ser	Leu	Asp	Pro	Thr	Phe	Thr	Ile
			1440						1445						

-continued

1475	1480	1485
Gly Arg Thr Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro 1490	1495	1500
Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys 1505	1510	1515 1520
Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr 1525	1530	1535
Val Arg Leu Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln 1540	1545	1550
Asp His Leu Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile 1555	1560	1565
Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Phe Pro 1570	1575	1580
Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro 1585	1590	1595 1600
Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro 1605	1610	1615
Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln 1620	1625	1630
Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Thr Cys 1635	1640	1645
Met Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly 1650	1655	1660
Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val 1665	1670	1675 1680
Val Ile Val Gly Arg Ile Val Leu Ser Gly Lys Pro Ala Ile Ile Pro 1685	1690	1695
Asp Arg Glu Val Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ser 1700	1705	1710
Gln His Leu Pro Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe 1715	1720	1725
Lys Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Ser Arg His Ala Glu 1730	1735	1740
Val Ile Thr Pro Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Val Phe 1745	1750	1755 1760
Trp Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala 1765	1770	1775
Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala 1780	1785	1790
Phe Thr Ala Ala Val Thr Ser Pro Leu Thr Thr Gly Gln Thr Leu Leu 1795	1800	1805
Phe Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro Gly 1810	1815	1820
Ala Ala Thr Ala Phe Val Gly Ala Gly Leu Ala Gly Ala Ala Ile Gly 1825	1830	1835 1840
Ser Val Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly 1845	1850	1855
Ala Gly Val Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu 1860	1865	1870
Val Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser 1875	1880	1885
Pro Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg 1890	1895	1900

-continued

His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile
 1905 1910 1915 1920
 Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro
 1925 1930 1935
 Glu Ser Asp Ala Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu Thr
 1940 1945 1950
 Val Thr Gln Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys
 1955 1960 1965
 Thr Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile
 1970 1975 1980
 Cys Glu Val Leu Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu Met
 1985 1990 1995 2000
 Pro Gln Leu Pro Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr Arg
 2005 2010 2015
 Gly Val Trp Arg Gly Asp Gly Ile Met His Thr Arg Cys His Cys Gly
 2020 2025 2030
 Ala Glu Ile Thr Gly His Val Lys Asn Gly Thr Met Arg Ile Val Gly
 2035 2040 2045
 Pro Arg Thr Cys Arg Asn Met Trp Ser Gly Thr Phe Pro Ile Asn Ala
 2050 2055 2060
 Tyr Thr Thr Gly Pro Cys Thr Pro Leu Pro Ala Pro Asn Tyr Lys Phe
 2065 2070 2075 2080
 Ala Leu Trp Arg Val Ser Ala Glu Glu Tyr Val Glu Ile Arg Arg Val
 2085 2090 2095
 Gly Asp Phe His Tyr Val Ser Gly Met Thr Thr Asp Asn Leu Lys Cys
 2100 2105 2110
 Pro Cys Gln Ile Pro Ser Pro Glu Phe Phe Thr Glu Leu Asp Gly Val
 2115 2120 2125
 Arg Leu His Arg Phe Ala Pro Pro Cys Lys Pro Leu Leu Arg Glu Glu
 2130 2135 2140
 Val Ser Phe Arg Val Gly Leu His Glu Tyr Pro Val Gly Ser Gln Leu
 2145 2150 2155 2160
 Pro Cys Glu Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr
 2165 2170 2175
 Asp Pro Ser His Ile Thr Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg
 2180 2185 2190
 Gly Ser Pro Pro Ser Met Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala
 2195 2200 2205
 Pro Ser Leu Lys Ala Thr Cys Thr Ala Asn His Asp Ser Pro Asp Ala
 2210 2215 2220
 Glu Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn
 2225 2230 2235 2240
 Ile Thr Arg Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe
 2245 2250 2255
 Asp Pro Leu Val Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala
 2260 2265 2270
 Glu Ile Leu Arg Lys Ser Arg Arg Phe Ala Arg Ala Leu Pro Val Trp
 2275 2280 2285
 Ala Arg Pro Asp Tyr Asn Pro Pro Leu Val Glu Thr Trp Lys Lys Pro
 2290 2295 2300
 Asp Tyr Glu Pro Pro Val His Gly Cys Pro Leu Pro Pro Pro Arg
 2305 2310 2315 2320

-continued

2740	2745	2750
Ala Ala Ser Leu Arg	Ala Phe Thr Glu Ala Met	Thr Arg Tyr Ser Ala
2755	2760	2765
Pro Pro Gly Asp Pro	Pro Gln Pro Glu Tyr Asp	Leu Glu Leu Ile Thr
2770	2775	2780
Ser Cys Ser Ser Asn Val	Ser Val Ala His Asp	Gly Ala Gly Lys Arg
2785	2790	2800
Val Tyr Tyr Leu Thr Arg	Asp Pro Thr Thr Pro	Leu Ala Arg Ala Ala
2805	2810	2815
Trp Glu Thr Ala Arg His	Thr Pro Val Asn Ser	Trp Leu Gly Asn Ile
2820	2825	2830
Ile Met Phe Ala Pro Thr	Leu Trp Ala Arg Met	Ile Leu Met Thr His
2835	2840	2845
Phe Phe Ser Val Leu Ile	Ala Arg Asp Gln Leu Glu	Gln Ala Leu Asn
2850	2855	2860
Cys Glu Ile Tyr Gly Ala	Cys Tyr Ser Ile Glu	Pro Leu Asp Leu Pro
2865	2870	2875
Pro Ile Ile Gln Arg Leu	His Gly Leu Ser Ala	Phe Ser Leu His Ser
2885	2890	2895
Tyr Ser Pro Gly Glu Ile	Asn Arg Val Ala Ala	Cys Leu Arg Lys Leu
2900	2905	2910
Gly Val Pro Pro Leu Arg	Ala Trp Arg His Arg	Ala Arg Ser Val Arg
2915	2920	2925
Ala Arg Leu Leu Ser Arg	Gly Gly Arg Ala Ala	Ile Cys Gly Lys Tyr
2930	2935	2940
Leu Phe Asn Trp Ala Val	Arg Thr Lys Leu Lys	Leu Thr Pro Ile Ala
2945	2950	2955
Ala Ala Gly Arg Leu Asp	Leu Ser Gly Trp Phe	Thr Ala Gly Tyr Ser
2965	2970	2975
Gly Gly Asp Ile Tyr His	Ser Val Ser His Ala	Arg Pro Arg Trp Phe
2980	2985	2990
Trp Phe Cys Leu Leu Leu	Leu Ala Ala Gly Val	Gly Ile Tyr Leu Leu
2995	3000	3005
Pro Asn Arg		
3010		

<210> SEQ ID NO 2

<211> LENGTH: 9599

<212> TYPE: DNA

<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 2

gccagccccc tgatgggggc gacactccac catgaatcac tcccctgtga ggaactactg	60
tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgcgtgcag cctccaggac	120
ccccctccc gggagagcca tagtggtctg cggaaccggt gactacaccg gaattgccag	180
gacgacggg tctttcttg gataaaccg ctcaatgcct ggagatttgg gcgtgcccc	240
gcaagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtgttact gcctgatagg	300
gtgcttgcca gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac	360
ctcaaagaaa aaccaaacgt aacaccaacc gtcgccaca ggacgtcaag ttcccggtg	420
gcggtcagat cgttggtgga gtttacttgt tgccgcgcag gggccctaga ttgggtgtgc	480
gcgcgacgag gaagacttcc gagcgggtcgc aacctcgagg tagacgtcag cctatcccca	540

-continued

aggcacgtcg	gccccagggc	aggacctggg	ctcagcccgg	gtacccttgg	ccccctctatg	600
gcaatgaggg	ttgcgggtgg	gcgggatggc	tcctgtctcc	ccgtggctct	cgccctagct	660
ggggccccc	agacccccgg	cgtaggtcgc	gcaatttggg	taaggtcac	gataccctta	720
cgtgcggcct	cgcgcacctc	atgggggtaca	taccgctcgt	cggcgccccc	cttgaggcg	780
ctgccagggc	cctggcgcat	ggcgctcggg	ttctggaaga	cggcgtgaac	tatgcaacag	840
ggaaccttcc	tggttgctct	ttctctatct	tccttctggc	cctgctctct	tgccctgactg	900
tgcccgcttc	agcctaccaa	gtgcgcaatt	cctcggggct	ttaccatgtc	accaatgatt	960
gccctaactc	gagtattgtg	tacgaggcgg	ccgatgccat	cctgcacact	ccgggggtgtg	1020
tccttgcgt	tcgcgagggt	aacgcctcga	ggtgttgggt	ggcggtgacc	cccacgggtg	1080
ccaccaggg	cggcaaacctc	cccacaacgc	agcttcgacg	tcatatcgat	ctgcttgtcg	1140
ggagcggcc	cctctgctcg	gccctctacg	tgggggacct	gtcgggtct	gtctttcttg	1200
ttggtcaact	gtttaccttc	tctcccaggc	gccactggac	gacgcaagac	tgcaattgtt	1260
ctatctatcc	cggccatata	acgggtcatc	gcattggcatg	ggatatgatg	atgaaactggt	1320
cccctacggc	agcgttgggt	gtagctcagc	tgctccggat	cccacaagcc	atcatggaca	1380
tgatcgctgg	tgctcactgg	ggagtcctgg	cgggcatagc	gtatttctcc	atggtgggga	1440
actggggcaa	ggtcctggta	gtgctgctgc	tatttgccgg	cgtcgacgcg	gaaacccacg	1500
tcaccggggg	aaatgcgggc	cgcaccacgg	ctgggcttgt	tggtctcctt	acaccaggcg	1560
ccaagcagaa	catccaactg	atcaacacca	acggcagttg	gcacatcaat	agcacggcct	1620
tgaattgcaa	tgaaagcctt	aacaccggct	ggttagcagg	gctcttctat	caacacaaat	1680
tcaactcttc	aggctgtcct	gagaggttgg	ccagctgccg	acgccttacc	gattttgccc	1740
agggctgggg	tcctatcagt	tatgccaaag	gaagcggcct	cgcgcaacgc	cctactgct	1800
ggcactaccc	tccaagacct	tgtggcattg	tgcccgcaaa	gagcgtgtgt	ggcccggtat	1860
attgcttcac	tcacagcccc	gtggtgggtg	gaacgaccga	caggctcggc	gcgcctacct	1920
acagctgggg	tgcaaatgat	acggatgtct	tcgtccttaa	caacaccagg	ccaccgctgg	1980
gcaattgggt	cggttgtacc	tggtgaact	caactggatt	caccaaagtg	tcgggagcgc	2040
ccccctgtgt	catcggaggg	gtgggcaaca	acaccttgct	ctgccccact	gattgcttcc	2100
gcaaacatcc	ggaagccaca	tactctcggg	gcggtccgg	tccttgatt	acaccagggt	2160
gcatggtcga	ctaccggtat	aggctttggc	actatccttg	taccatcaat	tacaccatat	2220
tcaaagtcag	gatgtacgtg	ggaggggtcg	agcacaggct	ggaagcggcc	tgcaactgga	2280
cgcggggcga	acgctgtgat	ctggaagaca	gggacaggtc	cgaagtcagc	ccgttgctgc	2340
tgccaccac	acagtggcag	gtccttcctg	gttctttcac	gacctgccca	gccttgcca	2400
ccggcctcat	ccacctccac	cagaacattg	tggacgtgca	gtactgttac	ggggtagggt	2460
caagcatcgc	gtcctggggc	attaagtggt	agtacgtcgt	tctcctgttc	cttctgcttg	2520
cagacgcgcg	cgtctgctcc	tgcttggtga	tgatgttact	catatcccaa	gcggaggcgg	2580
ctttggagaa	cctcgttaata	ctcaatgcag	catccctggc	cgggacgcac	ggtcttgtgt	2640
ccttctcctg	gttcttctgc	tttgctgggt	atctgaaggg	taggtgggtg	cccgagcgg	2700
tctacgccct	ctacgggatg	tggcctctcc	tcctgctcct	gctggcgttg	cctcagcggg	2760
catacgcaact	ggacacggag	gtggccgcgt	cgtgtggcgg	cgttggtctt	gtcgggttaa	2820
tggcgctgac	tctgtgccca	tattacaagc	gctatatacag	ctggtgcatg	tgggtgcttc	2880
agtattttct	gaccagagta	gaagcgcaac	tgacgtgtg	ggttcccccc	ctcaacgtcc	2940

-continued

ggggggggcg	cgatgccgtc	atcttactca	tgtgtgtagt	acacccgacc	ctggatattg	3000
acatcaccaa	actactcctg	gcatctctcg	gacccctttg	gattcttcaa	gccagtttgc	3060
ttaaagtcce	ctacttcgtg	cgcgttcaag	gccttctccg	gatctgcgcg	ctagcgcgga	3120
agatagccgg	aggtcattac	gtgcaaatgg	ccatcatcaa	gttaggggcg	cttactggca	3180
cctatgtgta	taaccatctc	acccctcttc	gagactgggc	gcacaacggc	ctgcgagatc	3240
tggccgtggc	tgtggaacca	gtcgtcttct	cccgaatgga	gaccaagctc	atcacgtggg	3300
gggcagatac	cgccgcgtgc	ggtgacatca	tcaacggctt	gcccgctctc	gcccgtaggg	3360
gccaggagat	actgcttggg	ccagccgacg	gaatggctct	caaggggtgg	aggttgctgg	3420
cggccatcac	ggcgtagccc	cagcagacga	gaggcctcct	agggtgtata	atcaccagcc	3480
tgactggccg	ggacaaaaac	caagtggagg	gtgaggtcca	gatcgtgtca	actgctaccc	3540
aaaccttctc	ggcaactgtc	atcaatgggg	tatgctggac	tgtctaccac	ggggccggaa	3600
cgaggaccat	cgcacacccc	aagggtcctg	tcatccagat	gtataccaat	gtggaccaag	3660
accttgtggg	ctggcccgcct	cctcaagggt	cccgtcatt	gacaccctgt	acctgcggct	3720
cctcggacct	ttacctggtc	acgaggcacg	ccgatgtcat	tcccgctcgc	cggcgagggtg	3780
atagcagggg	tagcctgctt	tcgccccggc	ccatttctta	cttgaaaggc	tcctcggggg	3840
gtccgctggt	gtgccccgcg	ggacacgcgc	tgggcctatt	cagggccgcg	gtgtgcaccc	3900
gtggagtggc	taaagcgggt	gactttatcc	ctgtggagaa	cctagggaca	accatgagat	3960
ccccggtggt	cacggacaac	tcctctccac	cagcagtgcc	ccagagcttc	cagggtggccc	4020
acctgcacgc	tcccaccggc	agcggtaaga	gcaccaaggt	cccggtcgcg	tacgcagccc	4080
agggctacaa	ggtgttgggt	ctcaaccctc	ctgttgctgc	aacgtggggc	tttgtgtctt	4140
acatgtccaa	ggcccatggg	gttgatccta	atatcaggac	cggggtgaga	acaattacca	4200
ctggcagccc	catcacgtac	tccacctacg	gcaagttcct	tgccgacggc	gggtgctcag	4260
gagggtgctta	tgacataata	atttgtgacg	agtgccactc	cacggatgcc	acatccatct	4320
tgggcacgcg	cactgtcctt	gaccaagcag	agactgcggg	ggcgagactg	gttgtgctcg	4380
ccactgctac	ccctccgggc	tccgtcactg	tgtcccatcc	taacatcgag	gaggttgctc	4440
tgtccaccac	cggagagatc	cccttttacg	gcaaggctat	ccccctcgag	gtgatcaagg	4500
ggggaagaca	tctcatcttc	tgccactcaa	agaagaagtg	cgacgagctc	gccgcgaagc	4560
tggtcgcatt	gggcataaat	gccgtggcct	actaccgcgg	tcttgacgtg	tctgtcatcc	4620
cgaaccagcg	cgatgttgtc	gtcgtgtcga	ccgatgctct	catgactggc	tttaccggcg	4680
acttcgactc	tgtgatagac	tgcaacacgt	gtgtcactca	gacagtcgat	ttcagccttg	4740
accctacctt	taccattgag	acaaccacgc	tccccaggga	tgtgtctctc	aggactcaac	4800
gccggggcag	gactggcagg	gggaagccag	gcacttatag	atttgtggca	ccgggggagc	4860
gcccctccgg	catgttcgac	tcgtccgtcc	tctgtgagtg	ctatgacgcg	ggctgtgctt	4920
ggtagtagct	cacgcccggc	gagactacag	ttaggctacg	agcgtacatg	aacaccccg	4980
ggcttcccg	gtgccaggac	catcttgaat	tttgggaggg	cgtctttacg	ggcctcactc	5040
atatagatgc	ccacttttta	tcccagacaa	agcagagtgg	ggagaacttt	ccttacctgg	5100
tagcgtacca	agccaccgtg	tcgcgtaggg	ctcaagcccc	tcccccatcg	tgggaccaga	5160
tgtggaagtg	tttgatccgc	cttaaaccca	ccctccatgg	gccaacaccc	ctgctataca	5220
gactgggcgc	tgttcagaat	gaagtcaccc	tgacgcaccc	aatcaccaaa	tacatcatga	5280

-continued

catgcatgtc	ggccgacctg	gaggctgtca	cgagcacctg	gggtgctggt	ggcggcgctc	5340
tggctgctct	ggccgcgtat	tgcctgtcaa	caggctgcgt	ggtcatagtg	ggcaggatcg	5400
tcttgctccg	gaagccggca	attataacctg	acagggaggt	tctctaccag	gagttcgatg	5460
agatggaaga	gtgctctcag	cacttacctg	acatcgagca	agggatgatg	ctcgctgagc	5520
agttcaagca	gaaggccctc	ggcctcctgc	agaccgcgtc	ccgccatgca	gaggttatca	5580
ccccctgctg	ccagaccaac	tggcagaaac	tcgaggtctt	ttgggcgaag	cacatgtgga	5640
atttcacatg	tgggatacaa	tacttgccgg	gcctgtcaac	gctgcctggt	aaccccgcca	5700
ttgcttcatt	gatggctttt	acagctgccg	tcaccagccc	actaaccact	ggccaaaccc	5760
tcctcttcaa	catattgggg	gggtgggtgg	ctgccagct	cgcgcgcctc	ggcgccgcta	5820
ctgcctttgt	gggtgctggc	ctagctggcg	ccgccatcgg	cagcgttgga	ctggggaagg	5880
tcctcgtgga	cattcttgca	gggtatggcg	cgggcgtggc	gggagctctt	gtagcattca	5940
agatcatgag	cggtagggtc	ccctccacgg	aggacctggt	caatctgctg	cccgccatcc	6000
tctgcctcgg	agcccttgta	gtcgggtgtg	tctgcgcagc	aatactgcgc	cggcacgttg	6060
gcccggggca	gggggcagtg	caatggatga	accggctaata	agccttcgcc	tcccggggga	6120
accatgtttc	ccccacgcac	tacgtgccgg	agagcgatgc	agccgcccgc	gtcactgcca	6180
tactcagcag	cctcactgta	accagctccc	tgaggcgact	gcatcagtgg	ataagctcgg	6240
agtgatacc	tccatgctcc	ggttcctggc	taaggacat	ctgggactgg	atatgcgagg	6300
tgctgagcga	ctttaagacc	tggctgaaag	ccaagctcat	gccacaactg	cctgggattc	6360
cctttgtgtc	ctgccagcgc	gggtataggg	gggtctggcg	aggagacggc	attatgcaca	6420
ctcgctgcca	ctgtggagct	gagatcactg	gacatgtcaa	aaacgggacg	atgaggatcg	6480
tcggtcctag	gacctgcagg	aacatgtgga	gtgggacgtt	ccccattaac	gcctacacca	6540
cgggcccctg	tactcccctt	cctgcgccga	actataagtt	cgcgctgtgg	aggggtgtctg	6600
cagaggaata	cgtggagata	aggcgggtgg	gggacttcca	ctacgtatcg	ggtatgacta	6660
ctgacaatct	taaatgcccg	tgccagatcc	catcgcccga	atttttcaca	gaattggacg	6720
gggtgcgctt	acacaggttt	gcgccccctt	gcaagccctt	gctgcgggag	gaggtatcat	6780
tcagagtagg	actccacgag	taccgggtgg	ggtcgcaatt	accttgcgag	cccgaaccgg	6840
acgtagccgt	gttagcgtcc	atgctcactg	atccctccca	tataacagca	gaggcggccg	6900
ggagaagggt	ggcgagaggg	tcaccccctt	ctatggccag	ctcctcggtt	agccagctgt	6960
ccgctccatc	tctcaaggca	acttgacccg	ccaacatga	ctcccctgac	gccgagctca	7020
tagaggctaa	cctcctgtgg	aggcaggaga	tgggcggcaa	catcaccagg	gttgagtcat	7080
agaacaaagt	ggtgattctg	gactccttcg	atccgcttgt	ggcagaggag	gatgagcggg	7140
aggtctccgt	acctgcagaa	attctgcgga	agtctcgag	attcgcccgg	gcctgccccg	7200
tctgggcgcg	gcgggactac	aacccccccg	tagtagagac	gtggaaaaag	cctgactacg	7260
aaccacctgt	ggtccatggc	tgcccgtac	cacctccacg	gtcccctcct	gtgctccgc	7320
ctcggaaaaa	gcgtacgggtg	gtcctcaccg	aatcaaccct	atctactgcc	ttggccgagc	7380
ttgccaccaa	aagttttggc	agctcctcaa	cttcgggcat	tacgggcgac	aatacgacaa	7440
catectctga	gcccccccct	tctgggtgcc	cccccgactc	cagcgttgag	tcctattctt	7500
ccatgcccc	cctggagggg	gagcctgggg	atccggatct	cagcgacggg	tcattggtcga	7560
cggtcagtag	tggggccgac	acggaagatg	tcgtgtgctg	ctcaatgtct	tattcctgga	7620
caggcgcaact	cgtcaccccc	tgcgctgcgg	aagaacaaaa	actgcccctc	aacgcactga	7680

-continued

```

gcaactcggt gctacccat cacaatctgg tgtattccac cacttcacgc agtgcttgcc 7740
aaaggcagaa gaaagtcaca tttagacagac tgcaagttct ggacagccat taccaggacg 7800
tgctcaagga ggtcaaagca gcggcgtcaa aagtgaaggc taacttgeta tccgtagagg 7860
aagcttgacg cctgacgccc ccacattcag ccaaatccaa gtttggctat ggggcaaaag 7920
acgtccgttg ccatgccaga aaggccgtag ccacatcaa ctccgtgtgg aaagaccttc 7980
tggaagacag tgtaacacca atagacacta ccatcatggc caagaacgag gttttctgcg 8040
ttcagcctga gaaggggggt cgtaagccag ctctgtctcat cgtgttcccc gacctgggcy 8100
tgcgcggtgt cgagaagatg gccctgtacg acgtgggttag caagctcccc ctggccgtga 8160
tggaagctc ctacggatcc caatactcac caggacagcg ggttgaattc ctctgcaag 8220
cgtggaagtc caagaagacc ccgatggggt tctctgatga taccgctgt tttgactcca 8280
cagtcactga gagcgacatc cgtacggagg aggcatttta ccaatgttgt gacctggacc 8340
cccaagcccc cgtggccatc aagtcctca ctgagaggct ttatgttggg ggccctctta 8400
ccaattcaag gggggaaaac tgcggctacc gcagggtccg cgcgagcggc gtactgacaa 8460
ctagctgtgg taacaccctc acttgctaca tcaaggcccc ggagcctgt cgagccgag 8520
ggctccagga ctgcaccatg ctctgtgtgt gcgacgactt agtcgttatc tgtgaaagt 8580
cgggggtcca ggaggacgcy gcgagcctga gagccttcac ggaggctatg accaggtact 8640
ccgccccccc cggggacccc ccacaaccag aatacgactt ggagcttata acatcatgct 8700
cctccaacgt gtcagtcgcc caccagcgcy ctggaaagag ggtctactac cttaccctg 8760
accctacaac cccctcgcy agagccgcy gggagacagc aagacacact ccagtcaatt 8820
cctggctagg caacataatc atgtttgccc ccacactgtg ggcgaggatg atactgatga 8880
cccatttctt tagcgtctc atagccagg atcagcttga acaggctctt aactgtgaga 8940
tctacggagc ctgctactcc atagaaccac tggatctacc tccaatcatt caaagactcc 9000
atggcctcag cgcattttca ctccacagtt actctccagg tgaatcaat aggggtggcc 9060
catgcctcag aaaacttggy gtcccgcct tgcgagcttg gagacaccgg gcccgagcg 9120
tccgctag gttctgtcc agaggaggca gggctgcat atgtggcaag tacctctca 9180
actgggcagt aagaacaaag ctcaactca ctccaatagc gccgctggc cggtggact 9240
tgtccggtt gttcacggct ggctacagcy ggggagacat ttatcacagc gtgtctcatg 9300
cccgccccg ctggttctg ttttgctac tctgtctgc tgcaggggta ggcactacc 9360
tctccccc aa cgatgaagg ttggggtaaa cactccggcc tcttaagcca tttctgttt 9420
ttttttttt ttttttttt tttttcttt ttttttctt tctttctct ctttttttc 9480
ttttttttt ctttttttaa tgggtgctcc atcttagccc tagtcacggc tagctgtgaa 9540
aggtccgtga gccgcatgac tgcagagagt gctgatactg gcctctctgc agatcatgt 9599

```

```

<210> SEQ ID NO 3
<211> LENGTH: 3010
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

```

```

<400> SEQUENCE: 3

```

```

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn
 1             5             10             15
Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly
          20             25             30

```

-continued

Gly	Val	Tyr	Leu	Leu	Pro	Arg	Arg	Gly	Pro	Arg	Leu	Gly	Val	Arg	Ala
	35						40					45			
Thr	Arg	Lys	Ala	Ser	Glu	Arg	Ser	Gln	Pro	Arg	Gly	Arg	Arg	Gln	Pro
	50					55					60				
Ile	Pro	Lys	Ala	Arg	Arg	Pro	Glu	Gly	Arg	Ala	Trp	Ala	Gln	Pro	Gly
	65				70					75					80
Tyr	Pro	Trp	Pro	Leu	Tyr	Gly	Asn	Glu	Gly	Leu	Gly	Trp	Ala	Gly	Trp
				85					90					95	
Leu	Leu	Ser	Pro	Arg	Gly	Ser	Arg	Pro	Ser	Trp	Gly	Pro	Thr	Asp	Pro
			100					105					110		
Arg	Arg	Arg	Ser	Arg	Asn	Leu	Gly	Lys	Val	Ile	Asp	Thr	Leu	Thr	Cys
			115				120				125				
Gly	Phe	Ala	Asp	Leu	Met	Gly	Tyr	Ile	Pro	Leu	Val	Gly	Ala	Pro	Leu
	130					135					140				
Gly	Gly	Ala	Ala	Arg	Ala	Leu	Ala	His	Gly	Val	Arg	Val	Leu	Glu	Asp
	145				150					155					160
Gly	Val	Asn	Tyr	Ala	Thr	Gly	Asn	Leu	Pro	Gly	Cys	Ser	Phe	Ser	Ile
				165					170					175	
Phe	Leu	Leu	Ala	Leu	Leu	Ser	Cys	Leu	Thr	Ile	Pro	Ala	Ser	Ala	Tyr
			180					185					190		
Glu	Val	Arg	Asn	Val	Ser	Gly	Ile	Tyr	His	Val	Thr	Asn	Asp	Cys	Ser
		195					200					205			
Asn	Ser	Ser	Ile	Val	Tyr	Glu	Ala	Ala	Asp	Val	Ile	Met	His	Thr	Pro
		210				215					220				
Gly	Cys	Val	Pro	Cys	Val	Gln	Glu	Gly	Asn	Ser	Ser	Arg	Cys	Trp	Val
	225				230					235					240
Ala	Leu	Thr	Pro	Thr	Leu	Ala	Ala	Arg	Asn	Ala	Ser	Val	Pro	Thr	Thr
				245					250					255	
Thr	Ile	Arg	Arg	His	Val	Asp	Leu	Leu	Val	Gly	Thr	Ala	Ala	Phe	Cys
			260				265						270		
Ser	Ala	Met	Tyr	Val	Gly	Asp	Leu	Cys	Gly	Ser	Ile	Phe	Leu	Val	Ser
		275				280						285			
Gln	Leu	Phe	Thr	Phe	Ser	Pro	Arg	Arg	His	Glu	Thr	Val	Gln	Asp	Cys
	290					295					300				
Asn	Cys	Ser	Ile	Tyr	Pro	Gly	His	Val	Ser	Gly	His	Arg	Met	Ala	Trp
	305				310					315					320
Asp	Met	Met	Met	Asn	Trp	Ser	Pro	Thr	Thr	Ala	Leu	Val	Val	Ser	Gln
				325					330					335	
Leu	Leu	Arg	Ile	Pro	Gln	Ala	Val	Val	Asp	Met	Val	Ala	Gly	Ala	His
			340					345					350		
Trp	Gly	Val	Leu	Ala	Gly	Leu	Ala	Tyr	Tyr	Ser	Met	Val	Gly	Asn	Trp
		355				360						365			
Ala	Lys	Val	Leu	Ile	Val	Ala	Leu	Leu	Phe	Ala	Gly	Val	Asp	Gly	Glu
	370					375					380				
Thr	His	Thr	Thr	Gly	Arg	Val	Ala	Gly	His	Thr	Thr	Ser	Gly	Phe	Thr
	385				390					395					400
Ser	Leu	Phe	Ser	Ser	Gly	Ala	Ser	Gln	Lys	Ile	Gln	Leu	Val	Asn	Thr
			405						410					415	
Asn	Gly	Ser	Trp	His	Ile	Asn	Arg	Thr	Ala	Leu	Asn	Cys	Asn	Asp	Ser
			420					425					430		
Leu	Gln	Thr	Gly	Phe	Phe	Ala	Ala	Leu	Phe	Tyr	Ala	His	Lys	Phe	Asn
	435						440					445			
Ser	Ser	Gly	Cys	Pro	Glu	Arg	Met	Ala	Ser	Cys	Arg	Pro	Ile	Asp	Trp

450					455					460					
Phe	Ala	Gln	Gly	Trp	Gly	Pro	Ile	Thr	Tyr	Thr	Lys	Pro	Asn	Ser	Ser
465					470					475					480
Asp	Gln	Arg	Pro	Tyr	Cys	Trp	His	Tyr	Ala	Pro	Arg	Pro	Cys	Gly	Val
				485					490					495	
Val	Pro	Ala	Ser	Gln	Val	Cys	Gly	Pro	Val	Tyr	Cys	Phe	Thr	Pro	Ser
			500					505					510		
Pro	Val	Val	Val	Gly	Thr	Thr	Asp	Arg	Ser	Gly	Val	Pro	Thr	Tyr	Ser
			515				520					525			
Trp	Gly	Glu	Asn	Glu	Thr	Asp	Val	Met	Leu	Leu	Asn	Asn	Thr	Arg	Pro
	530					535					540				
Pro	Gln	Gly	Asn	Trp	Phe	Gly	Cys	Thr	Trp	Met	Asn	Ser	Thr	Gly	Phe
545					550					555					560
Thr	Lys	Thr	Cys	Gly	Gly	Pro	Pro	Cys	Asn	Ile	Gly	Gly	Val	Gly	Asn
				565					570					575	
Arg	Thr	Leu	Ile	Cys	Pro	Thr	Asp	Cys	Phe	Arg	Lys	His	Pro	Glu	Ala
			580					585					590		
Thr	Tyr	Thr	Lys	Cys	Gly	Ser	Gly	Pro	Trp	Leu	Thr	Pro	Arg	Cys	Leu
			595				600					605			
Val	Asp	Tyr	Pro	Tyr	Arg	Leu	Trp	His	Tyr	Pro	Cys	Thr	Leu	Asn	Phe
	610					615					620				
Ser	Ile	Phe	Lys	Val	Arg	Met	Tyr	Val	Gly	Gly	Val	Glu	His	Arg	Leu
625					630					635					640
Asn	Ala	Ala	Cys	Asn	Trp	Thr	Arg	Gly	Glu	Arg	Cys	Asn	Leu	Glu	Asp
				645					650					655	
Arg	Asp	Arg	Ser	Glu	Leu	Ser	Pro	Leu	Leu	Leu	Ser	Thr	Thr	Glu	Trp
			660					665					670		
Gln	Ile	Leu	Pro	Cys	Ala	Phe	Thr	Thr	Leu	Pro	Ala	Leu	Ser	Thr	Gly
			675				680					685			
Leu	Ile	His	Leu	His	Gln	Asn	Ile	Val	Asp	Val	Gln	Tyr	Leu	Tyr	Gly
	690					695					700				
Val	Gly	Ser	Ala	Phe	Val	Ser	Phe	Ala	Ile	Lys	Trp	Glu	Tyr	Ile	Leu
705					710					715					720
Leu	Leu	Phe	Leu	Leu	Leu	Ala	Asp	Ala	Arg	Val	Cys	Ala	Cys	Leu	Trp
				725					730				735		
Met	Met	Leu	Leu	Ile	Ala	Gln	Ala	Glu	Ala	Ala	Leu	Glu	Asn	Leu	Val
			740					745					750		
Val	Leu	Asn	Ala	Ala	Ser	Val	Ala	Gly	Ala	His	Gly	Ile	Leu	Ser	Phe
			755				760					765			
Leu	Val	Phe	Phe	Cys	Ala	Ala	Trp	Tyr	Ile	Lys	Gly	Arg	Leu	Ala	Pro
	770					775					780				
Gly	Ala	Ala	Tyr	Ala	Phe	Tyr	Gly	Val	Trp	Pro	Leu	Leu	Leu	Leu	Leu
785					790					795					800
Leu	Ala	Leu	Pro	Pro	Arg	Ala	Tyr	Ala	Leu	Asp	Arg	Glu	Met	Ala	Ala
				805					8						

-continued

His Pro Glu Leu Ile Phe Asp Ile Thr Lys Leu Leu Leu Ala Ile Leu
 885 890 895
 Gly Pro Leu Met Val Leu Gln Ala Gly Ile Thr Arg Val Pro Tyr Phe
 900 905 910
 Val Arg Ala Gln Gly Leu Ile Arg Ala Cys Met Leu Val Arg Lys Val
 915 920 925
 Ala Gly Gly His Tyr Val Gln Met Val Phe Met Lys Leu Gly Ala Leu
 930 935 940
 Thr Gly Thr Tyr Val Tyr Asn His Leu Thr Pro Leu Arg Asp Trp Ala
 945 950 955 960
 His Ala Gly Leu Arg Asp Leu Ala Val Ala Val Glu Pro Val Val Phe
 965 970 975
 Ser Ala Met Glu Thr Lys Val Ile Thr Trp Gly Ala Asp Thr Ala Ala
 980 985 990
 Cys Gly Asp Ile Ile Leu Gly Leu Pro Val Ser Ala Arg Arg Gly Lys
 995 1000 1005
 Glu Ile Phe Leu Gly Pro Ala Asp Ser Leu Glu Gly Gln Gly Trp Arg
 1010 1015 1020
 Leu Leu Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Val Leu
 1025 1030 1035 1040
 Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu
 1045 1050 1055
 Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr
 1060 1065 1070
 Cys Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys
 1075 1080 1085
 Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val
 1090 1095 1100
 Asp Leu Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Met
 1105 1110 1115 1120
 Thr Pro Cys Ser Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His
 1125 1130 1135
 Ala Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu
 1140 1145 1150
 Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro
 1155 1160 1165
 Leu Leu Cys Pro Ser Gly His Val Val Gly Val Phe Arg Ala Ala Val
 1170 1175 1180
 Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Ser
 1185 1190 1195 1200
 Met Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Thr Pro
 1205 1210 1215
 Pro Ala Val Pro Gln Thr Phe Gln Val Ala His Leu His Ala Pro Thr
 1220 1225 1230
 Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly
 1235 1240 1245
 Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe
 1250 1255 1260
 Gly Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr
 1265 1270 1275 1280
 Gly Val Arg Thr Ile Thr Thr Gly Gly Ser Ile Thr Tyr Ser Thr Tyr
 1285 1290 1295

-continued

Gly	Lys	Phe	Leu	Ala	Asp	Gly	Gly	Cys	Ser	Gly	Gly	Ala	Tyr	Asp	Ile
			1300					1305						1310	
Ile	Ile	Cys	Asp	Glu	Cys	His	Ser	Thr	Asp	Ser	Thr	Thr	Ile	Leu	Gly
		1315					1320					1325			
Ile	Gly	Thr	Val	Leu	Asp	Gln	Ala	Glu	Thr	Ala	Gly	Ala	Arg	Leu	Val
	1330				1335					1340					
Val	Leu	Ala	Thr	Ala	Thr	Pro	Pro	Gly	Ser	Val	Thr	Val	Pro	His	Pro
1345			1350					1355					1360		
Asn	Ile	Glu	Glu	Ile	Gly	Leu	Ser	Asn	Asn	Gly	Glu	Ile	Pro	Phe	Tyr
			1365				1370						1375		
Gly	Lys	Ala	Ile	Pro	Ile	Glu	Ala	Ile	Lys	Gly	Gly	Arg	His	Leu	Ile
		1380					1385					1390			
Phe	Cys	His	Ser	Lys	Lys	Lys	Cys	Asp	Glu	Leu	Ala	Ala	Lys	Leu	Thr
	1395					1400					1405				
Gly	Leu	Gly	Leu	Asn	Ala	Val	Ala	Tyr	Tyr	Arg	Gly	Leu	Asp	Val	Ser
	1410				1415					1420					
Val	Ile	Pro	Pro	Ile	Gly	Asp	Val	Val	Val	Val	Ala	Thr	Asp	Ala	Leu
1425				1430				1435						1440	
Met	Thr	Gly	Phe	Thr	Gly	Asp	Phe	Asp	Ser	Val	Ile	Asp	Cys	Asn	Thr
			1445				1450					1455			
Cys	Val	Thr	Gln	Thr	Val	Asp	Phe	Ser	Leu	Asp	Pro	Thr	Phe	Thr	Ile
		1460				1465					1470				
Glu	Thr	Thr	Thr	Val	Pro	Gln	Asp	Ala	Val	Ser	Arg	Ser	Gln	Arg	Arg
	1475				1480						1485				
Gly	Arg	Thr	Gly	Arg	Gly	Arg	Ser	Gly	Ile	Tyr	Arg	Phe	Val	Thr	Pro
	1490				1495				1500						
Gly	Glu	Arg	Pro	Ser	Gly	Met	Phe	Asp	Ser	Ser	Val	Leu	Cys	Glu	Cys
1505				1510				1515					1520		
Tyr	Asp	Ala	Gly	Cys	Ala	Trp	Tyr	Glu	Leu	Thr	Pro	Ala	Glu	Thr	Ser
			1525				1530				1535				
Val	Arg	Leu	Arg	Ala	Tyr	Leu	Asn	Thr	Pro	Gly	Leu	Pro	Val	Cys	Gln
		1540				1545					1550				
Asp	His	Leu	Glu	Phe	Trp	Glu	Ser	Val	Phe	Thr	Gly	Leu	Thr	His	Ile
	1555				1560					1565					
Asp	Ala	His	Phe	Leu	Ser	Gln	Thr	Lys	Gln	Ala	Gly	Asp	Asn	Phe	Pro
	1570				1575				1580						
Tyr	Leu	Val	Ala	Tyr	Gln	Ala	Thr	Val	Cys	Ala	Arg	Ala	Gln	Ala	Pro
1585				1590				1595					1600		
Pro	Pro	Ser	Trp	Asp	Gln	Met	Trp	Lys	Cys	Leu	Ile	Arg	Leu	Lys	Pro
		1605				1610					1615				
Thr	Leu	His	Gly	Pro	Thr	Pro	Leu	Leu	Tyr	Arg	Leu	Gly	Ala	Val	Gln
		1620				1625					1630				
Asn	Glu	Val	Ile	Leu	Thr	His	Pro	Ile	Thr	Lys	Tyr	Ile	Met	Ala	Cys
	1635				1640						1645				
Met	Ser	Ala	Asp	Leu	Glu	Val	Val	Thr	Ser	Thr	Trp	Val	Leu	Val	Gly
	1650				1655					1660					
Gly	Val	Leu	Ala	Ala	Leu	Ala	Ala	Tyr	Cys	Leu	Thr	Thr	Gly	Ser	Val
1665				1670				1675					1680		
Val	Ile	Val	Gly	Arg	Ile	Ile	Leu	Ser	Gly	Lys	Pro	Ala	Val	Val	Pro
		1685				1690					1695				
Asp	Arg	Glu	Val	Leu	Tyr	Gln	Glu	Phe	Asp	Glu	Met	Glu	Glu	Cys	Ala
	1700					1705					1710				
Ser	Gln	Leu	Pro	Tyr	Ile	Glu	Gln	Gly	Met	Gln	Leu	Ala	Glu	Gln	Phe

-continued

1715	1720	1725
Lys Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu 1730 1735 1740		
Ala Ala Ala Pro Val Val Glu Ser Lys Trp Arg Ala Leu Glu Thr Phe 1745 1750 1755 1760		
Trp Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala 1765 1770 1775		
Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala 1780 1785 1790		
Phe Thr Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln Asn Thr Leu Leu 1795 1800 1805		
Phe Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser 1810 1815 1820		
Ala Ala Ser Ala Phe Val Gly Ala Gly Ile Ala Gly Ala Ala Val Gly 1825 1830 1835 1840		
Ser Ile Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly 1845 1850 1855		
Ala Gly Val Ala Gly Ala Leu Val Ala Phe Lys Val Met Ser Gly Glu 1860 1865 1870		
Val Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser 1875 1880 1885		
Pro Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg 1890 1895 1900		
His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile 1905 1910 1915 1920		
Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro 1925 1930 1935		
Glu Ser Asp Ala Ala Ala Arg Val Thr Gln Ile Leu Ser Ser Leu Thr 1940 1945 1950		
Ile Thr Gln Leu Leu Lys Arg Leu His Gln Trp Ile Asn Glu Asp Cys 1955 1960 1965		
Ser Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Val Trp Asp Trp Ile 1970 1975 1980		
Cys Thr Val Leu Thr Asp Phe Lys Thr Trp Leu Gln Ser Lys Leu Leu 1985 1990 1995 2000		
Pro Arg Leu Pro Gly Val Pro Phe Leu Ser Cys Gln Arg Gly Tyr Lys 2005 2010 2015		
Gly Val Trp Arg Gly Asp Gly Ile Met Gln Thr Thr Cys Pro Cys Gly 2020 2025 2030		
Ala Gln Ile Ala Gly His Val Lys Asn Gly Ser Met Arg Ile Val Gly 2035 2040 2045		
Pro Arg Thr Cys Ser Asn Thr Trp His Gly Thr Phe Pro Ile Asn Ala 2050 2055 2060		
Tyr Thr Thr Gly Pro Cys Thr Pro Ser Pro Ala Pro Asn Tyr Ser Arg 2065 2070 2075 2080		
Ala Leu Trp Arg Val Ala Ala Glu Glu Tyr Val Glu Val Thr Arg Val 2085 2090 2095		
Gly Asp Phe His Tyr Val Thr Gly Met Thr Thr Asp Asn Val Lys Cys 2100 2105 2110		
Pro Cys Gln Val Pro Ala Pro Glu Phe Phe Thr Glu Val Asp Gly Val 2115 2120 2125		
Arg Leu His Arg Tyr Ala Pro Ala Cys Lys Pro Leu Leu Arg Glu Asp 2130 2135 2140		

-continued

Val Thr Phe Gln Val Gly Leu Asn Gln Tyr Leu Val Gly Ser Gln Leu		
2145	2150	2155 2160
Pro Cys Glu Pro Glu Pro Asp Val Thr Val Leu Thr Ser Met Leu Thr		
	2165	2170 2175
Asp Pro Ser His Ile Thr Ala Glu Thr Ala Lys Arg Arg Leu Ala Arg		
	2180	2185 2190
Gly Ser Pro Pro Ser Leu Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala		
	2195	2200 2205
Pro Ser Leu Lys Ala Thr Cys Thr Thr His His Asp Ser Pro Asp Ala		
	2210	2215 2220
Asp Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn		
	2225	2230 2235 2240
Ile Thr Arg Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe		
	2245	2250 2255
Glu Pro Leu His Ala Glu Gly Asp Glu Arg Glu Ile Ser Val Ala Ala		
	2260	2265 2270
Glu Ile Leu Arg Lys Ser Arg Lys Phe Pro Ser Ala Leu Pro Ile Trp		
	2275	2280 2285
Ala Arg Pro Asp Tyr Asn Pro Pro Leu Leu Glu Ser Trp Lys Asp Pro		
	2290	2295 2300
Asp Tyr Val Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Thr Lys		
	2305	2310 2315 2320
Ala Pro Pro Ile Pro Pro Pro Arg Arg Lys Arg Thr Val Val Leu Thr		
	2325	2330 2335
Glu Ser Asn Val Ser Ser Ala Leu Ala Glu Leu Ala Thr Lys Thr Phe		
	2340	2345 2350
Gly Ser Ser Gly Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Leu		
	2355	2360 2365
Pro Asp Leu Ala Ser Asp Asp Gly Asp Lys Gly Ser Asp Val Glu Ser		
	2370	2375 2380
Tyr Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu		
	2385	2390 2395 2400
Ser Asp Gly Ser Trp Ser Thr Val Ser Glu Glu Ala Ser Glu Asp Val		
	2405	2410 2415
Val Cys Cys Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro		
	2420	2425 2430
Cys Ala Ala Glu Glu Ser Lys Leu Pro Ile Asn Pro Leu Ser Asn Ser		
	2435	2440 2445
Leu Leu Arg His His Asn Met Val Tyr Ala Thr Thr Ser Arg Ser Ala		
	2450	2455 2460
Ser Leu Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp		
	2465	2470 2475 2480
Asp His Tyr Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr		
	2485	2490 2495
Val Lys Ala Lys Leu Leu Ser Ile Glu Glu Ala Cys Lys Leu Thr Pro		
	2500	2505 2510
Pro His Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg		
	2515	2520 2525
Asn Leu Ser Ser Arg Ala Val Asn His Ile Arg Ser Val Trp Glu Asp		
	2530	2535 2540
Leu Leu Glu Asp Thr Glu Thr Pro Ile Asp Thr Thr Ile Met Ala Lys		
	2545	2550 2555 2560

-continued

Ser	Glu	Val	Phe	Cys	Val	Gln	Pro	Glu	Lys	Gly	Gly	Arg	Lys	Pro	Ala	
				2565					2570					2575		
Arg	Leu	Ile	Val	Phe	Pro	Asp	Leu	Gly	Val	Arg	Val	Cys	Glu	Lys	Met	
			2580					2585					2590			
Ala	Leu	Tyr	Asp	Val	Val	Ser	Thr	Leu	Pro	Gln	Ala	Val	Met	Gly	Ser	
		2595					2600					2605				
Ser	Tyr	Gly	Phe	Gln	Tyr	Ser	Pro	Lys	Gln	Arg	Val	Glu	Phe	Leu	Val	
	2610				2615					2620						
Asn	Thr	Trp	Lys	Ser	Lys	Lys	Cys	Pro	Met	Gly	Phe	Ser	Tyr	Asp	Thr	
2625				2630						2635				2640		
Arg	Cys	Phe	Asp	Ser	Thr	Val	Thr	Glu	Ser	Asp	Ile	Arg	Val	Glu	Glu	
		2645						2650					2655			
Ser	Ile	Tyr	Gln	Cys	Cys	Asp	Leu	Ala	Pro	Glu	Ala	Arg	Gln	Ala	Ile	
	2660						2665					2670				
Arg	Ser	Leu	Thr	Glu	Arg	Leu	Tyr	Ile	Gly	Gly	Pro	Leu	Thr	Asn	Ser	
	2675				2680						2685					
Lys	Gly	Gln	Asn	Cys	Gly	Tyr	Arg	Arg	Cys	Arg	Ala	Ser	Gly	Val	Leu	
	2690			2695					2700							
Thr	Thr	Ser	Cys	Gly	Asn	Thr	Leu	Thr	Cys	Tyr	Leu	Lys	Ala	Thr	Ala	
2705			2710					2715					2720			
Ala	Cys	Arg	Ala	Ala	Lys	Leu	Gln	Asp	Cys	Thr	Met	Leu	Val	Asn	Gly	
		2725					2730				2735					
Asp	Asp	Leu	Val	Val	Ile	Cys	Glu	Ser	Ala	Gly	Thr	Gln	Glu	Asp	Ala	
	2740					2745					2750					
Ala	Ala	Leu	Arg	Ala	Phe	Thr	Glu	Ala	Met	Thr	Arg	Tyr	Ser	Ala	Pro	
	2755			2760						2765						
Pro	Gly	Asp	Pro	Pro	Gln	Pro	Glu	Tyr	Asp	Leu	Glu	Leu	Ile	Thr	Ser	
	2770			2775					2780							
Cys	Ser	Ser	Asn	Val	Ser	Val	Ala	His	Asp	Ala	Ser	Gly	Lys	Arg	Val	
2785			2790				2795					2800				
Tyr	Tyr	Leu	Thr	Arg	Asp	Pro	Thr	Thr	Pro	Leu	Ala	Arg	Ala	Ala	Trp	
		2805				2810				2815						
Glu	Thr	Ala	Arg	His	Thr	Pro	Ile	Asn	Ser	Trp	Leu	Gly	Asn	Ile	Ile	
	2820				2825				2830							
Met	Tyr	Ala	Pro	Thr	Leu	Trp	Ala	Arg	Met	Ile	Leu	Met	Thr	His	Phe	
	2835			2840					2845							
Phe	Ser	Ile	Leu	Leu	Ala	Gln	Glu	Gln	Leu	Glu	Lys	Ala	Leu	Asp	Cys	
	2850			2855					2860							
Gln	Ile	Tyr	Gly	Ala	Cys	Tyr	Ser	Ile	Glu	Pro	Leu	Asp	Leu	Pro	Gln	
2865			2870				2875				2880					
Ile	Ile	Glu	Arg	Leu	His	Gly	Leu	Ser	Ala	Phe	Thr	Leu	His	Ser	Tyr	
	2885				2890				2895							
Ser	Pro	Gly	Glu	Ile	Asn	Arg	Val	Ala	Ser	Cys	Leu	Arg	Lys	Leu	Gly	
	2900				2905					2910						
Val	Pro	Pro	Leu	Arg	Thr	Trp	Arg	His	Arg	Ala	Arg	Ser	Val	Arg	Ala	
	2915				2920					2925						
Lys	Leu	Leu	Ser	Gln	Gly	Gly	Arg	Ala	Ala	Thr	Cys	Gly	Arg	Tyr	Leu	
	2930			2935					2940							
Phe	Asn	Trp	Ala	Val	Arg	Thr	Lys	Leu	Lys	Leu	Thr	Pro	Ile	Pro	Ala	
2945			2950				2955				2960					
Ala	Ser	Gln	Leu	Asp	Leu	Ser	Gly	Trp	Phe	Val	Ala	Gly	Tyr	Ser	Gly	
	2965				2970				2975							
Gly	Asp	Ile	Tyr	His	Ser	Leu	Ser	Arg	Ala	Arg	Pro	Arg	Trp	Phe	Pro	

-continued

2980	2985	2990	
Leu Cys Leu Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro			
2995	3000	3005	
Asn Arg			
3010			
 <210> SEQ ID NO 4			
<211> LENGTH: 9595			
<212> TYPE: DNA			
<213> ORGANISM: Hepatitis C virus			
 <400> SEQUENCE: 4			
gccagcccc	tgatggggc	gacactccac	catgaatcac tcccctgtga ggaactactg 60
tcttcacgca	gaaagcgtct	agccatggcg	ttagtatgag tgtcgtgcag cctccaggac 120
ccccctccc	gggagagcca	tagtgggtcg	cggaaaccgg gagtacaccg gaattgccag 180
gacgaccggg	tctttctctg	gatcaaccgg	ctcaatgcct ggagatttgg gcgtgcccc 240
gcgagactgc	tagccgagta	gtgttgggtc	gcgaaaggcc ttgtgttact gctgatagg 300
gtgcttgcca	gtgccccggg	aggtctcgta	gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa	aaccaaacgt	aacaccaacc	gccgccccaca ggacgtcaag ttcccgggcg 420
gtggtcagat	cgttggtgga	gtttacctgt	tgccgcgcag gggccccagg ttgggtgtgc 480
gcgcgactag	gaaggcttcc	gagcgggtcg	aacctcgtgg aaggcgacaa cctatcccaa 540
aggctcgccg	acccgagggc	agggcctggg	ctcagcccgg gtacccttgg cccctctatg 600
gcaatgaggg	cctgggggtg	gcaggatggc	tcctgtcacc ccgcggctcc cggcctagtt 660
ggggccccac	ggacccccgg	cgtaggtcgc	gtaacttggg taaggtcac gataccetta 720
catgcggctt	cgccgatctc	atgggggtaca	ttccgctcgt cggcgcccc ctagggggcg 780
ctgccagggc	cttggcacac	ggtgtccggg	ttctggagga cggcgtgaac tatgcaacag 840
ggaacttgcc	cggttgctct	ttctctatct	tcctcttggc tctgctgtcc tgtttgacca 900
tcccagcttc	cgcttatgaa	gtgcgcaacg	tgtccgggat ataccatgtc acgaacgact 960
gctccaaact	aagcatttgt	tatgaggcag	cggacgtgat catgcatact cccgggtgcg 1020
tgccctgtgt	tcaggagggt	aacagctccc	gttgetgggt agcgtcact cccacgctcg 1080
cggccaggaa	tgccagcgtc	cccactacga	caatacgacg ccacgtcgac ttgctcgttg 1140
ggacgggtgc	tttctgtctc	gctatgtacg	tggggggatct ctgcggatct attttcctcg 1200
tctcccagct	gttcaccttc	tcgcctcgcc	ggcatgagac agtgcaggac tgcaactgct 1260
caatctatcc	cggccatgta	tcaggtcacc	gcattggcttg ggatatgatg atgaactggt 1320
cacctacaac	agccctagtg	gtgtcgcagt	tgctccggat cccacaagct gtcgtggaca 1380
tggtggcggg	ggcccaactg	ggagtcctgg	cgggccttgc ctactattcc atggtaggga 1440
actgggctaa	ggttctgatt	gtggcgctac	tctttgcccg cgttgacggg gagaccacaa 1500
cgacggggag	ggtggccggc	cacaccacct	cggggttcac gtccttttc tcatctgggg 1560
cgtctcagaa	aatccagctt	gtgaatacca	acggcagctg gcacatcaac aggactgccc 1620
taaaattgcaa	tgactccctc	caaactgggt	tctttgcccg gctgttttac gcacacaagt 1680
tcaactcgtc	cgggtgcccc	gagcgcattg	ccagctgcgc cccattgac tggttcggcc 1740
aggggtgggg	ccccatcacc	tatactaagc	ctaacagctc ggatcagagg ccttattgct 1800
ggcattacgc	gcctcgaccg	tgtggtgtcg	taccgcgctc gcagggtgtg ggtccagtgt 1860
attgtttcac	cccaagccct	gttgtggtgg	ggaccaccga tcgttccggg gtcctacgt 1920

-continued

atagctgggg	ggagaatgag	acagacgtga	tgctcctcaa	caacacgcgt	ccgccacaag	1980
gcaactgggt	cggctgtaca	tggaatgaata	gtactgggtt	cactaagacg	tgcggaggtc	2040
ccccgtgtaa	catcgggggg	gtcggtaacc	gcaccttgat	ctgccccacg	gactgtctcc	2100
ggaagcacc	cagggtact	tacacaaaat	gtggctcggg	gccttggttg	acacctaggt	2160
gcctagtaga	ctaccatac	aggctttggc	actaccctcg	cactctcaat	ttttccatct	2220
ttaaggtag	gatgtatgtg	ggggcggtg	agcacaggct	caatgccgca	tgcaattgga	2280
ctcgaggaga	gcgctgtaac	ttggaggaca	gggataagtc	agaactcagc	ccgctgctgc	2340
tgctataaac	agagtggcag	atactgccct	gtgctttcac	cacctaccg	gctttatcca	2400
ctggtttgat	ccatctccat	cagaacatcg	tggacgtgca	atacctgtac	ggtgtaggg	2460
cagcgtttgt	ctcctttgca	atcaaatggg	agtacatcct	gttgcttttc	cttctcctgg	2520
cagacgcgcg	cgtgtgtgcc	tgcttggtga	tgatgctgct	gatagcccag	gctgaggccg	2580
ccttagagaa	cttgggtggc	ctcaatgcgg	cgcccggtgc	cggagcgcac	ggtattctct	2640
cctttcttgt	gttctcttgc	gccgcctgg	acattaaggg	caggctggct	cctggggcgg	2700
cgtatgcttt	ttatggcgta	tggccgctgc	tcctgctcct	actggcggtta	ccaccacgag	2760
cttacgcctt	ggaccgggag	atggctgcat	cgtgcggggg	tgcggttctt	gtaggctctg	2820
tattcttgac	cttggtacca	tactacaaag	tgtttctcac	taggctcata	tgggtgttac	2880
aatactttat	caccagagcc	gaggcgacac	tgcaagtgtg	gggtcccccc	ctcaacgttc	2940
ggggaggccg	cgatgccatc	atcctctca	cgtgtgcgg	tcattccagag	ttaatttttg	3000
acatcaccaa	actcctgtgc	gccatactcg	gcccgctcat	ggtgctccag	gctggcataa	3060
cgagagtgcc	gtacttcgtg	cgcgctcaag	ggctcattcg	tgcattgcag	ttagtgcgaa	3120
aagtgcggcg	gggtcattat	gtccaaatgg	tcttcatgaa	gctgggcgcg	ctgacaggta	3180
cgtacgttta	taaccatctt	acccactgc	gggactgggc	ccagcggggc	ctacgagacc	3240
ttgcgggtgc	ggttagagcc	gtcgtcttct	ccgccatgga	gaccaaggtc	atcacctggg	3300
gagcagacac	cgtgcgtgtg	ggggacatca	tcttgggtct	accgctctcc	gcccgaagg	3360
ggaaggagat	atttttggga	cgggtgata	gtctcgaagg	gcaagggtgg	cgactccttg	3420
cgcccatcac	ggcctactcc	caacaaacgc	ggggcgta	tggttgcatc	atcactagcc	3480
tcacaggccg	ggacaagaac	caggctgaag	gggaggttca	agtggtttct	accgcaaac	3540
aatctttcct	ggcgacctgc	atcaacggcg	tgtgtgggac	tgtctaccat	ggcgctggct	3600
cgaagacct	agcgggtcca	aaagggtcaa	tcacccaaat	gtacaccaat	gtagacctgg	3660
acctcgctcg	ctggcaggcg	cccccgggg	cgcgctccat	gacacctgc	agctgtggca	3720
gctcgacct	ttacttggtc	acgagacatg	ctgatgtcat	tccggtgcgc	cggcgaggcg	3780
acagcagggg	aagtctactc	tccccaggc	cgtctccta	cctgaaaggc	tcctcgggtg	3840
gtccattgct	ttgcccttcg	gggcacgtcg	tgggcgtctt	ccgggctgct	gtgtgcaccc	3900
ggggggtcgc	gaaggcggtg	gacttcatac	ccgttgagtc	tatggaaact	accatgcgg	3960
ctcgggtctt	cacagacaa	tcaaccccc	cggctgtacc	gcagacattc	caagtggcac	4020
atctgcacgc	tcctactggc	agcggcaaga	gcaccaaagt	gccggctgcg	tatgcagccc	4080
aagggtacaa	ggtgctcgtc	ctgaacccgt	ccgttgccgc	caccttaggg	tttggggcgt	4140
atatgtccaa	ggcacacggt	atcgacccta	acatcagaac	tggggtaagg	accattacca	4200
cgggcggctc	cattacgtac	tccacctatg	gcaagttcct	tgcgcagcgt	ggctgttctg	4260

-continued

gggggccta	tgacatcata	atatgtgatg	agtgcactc	aactgactcg	actaccatct	4320
tgggcatcgg	cacagtcctg	gaccaagcgg	agacggctgg	agcgcgctc	gtcgtgctcg	4380
ccaccgctac	acctccggga	tcggttacgg	tgccacaccc	caatatcgag	gaaataggcc	4440
tgtccaacaa	tggagagatc	cccttctatg	gcaaagccat	ccccattgag	gccatcaagg	4500
gggggaggca	tctcattttc	tgccattcca	agaagaaatg	tgacgagctc	gccgcaaagc	4560
tgacaggcct	cggactgaac	gctgtagcat	attaccgggg	ccttgatgtg	tccgtcatac	4620
cgcctatcgg	agacgtcgtt	gtcgtggcaa	cagacgtctc	aatgacgggt	ttaccggcgg	4680
atthtgactc	agtgatcgac	tgcaatacat	gtgtcaccca	gacagtcgac	ttcagcttgg	4740
atcccacctt	caccattgag	acgacgacgg	tgccccaaga	cgcggtgtcg	cgctcgcaac	4800
ggcgaggtag	aactggcagg	ggtaggagtg	gcactctacag	gtttgtgact	ccaggagaaac	4860
ggccctcggg	catgttcgat	tcttcgggcc	tgtgtgagtg	ctatgacgcg	ggctgtgctt	4920
gggtatgagct	cacgcccgtc	gagacctcgg	ttaggttcgg	ggcttaccta	aatacaccag	4980
gggtgcccg	ctgccaggac	catctggagt	tctgggagag	cgtcttcaca	ggcctcacc	5040
acatagatgc	ccacttcctg	tcccagacta	aacaggcagg	agacaacttt	ccttacctgg	5100
tggcatatca	agctacatg	tgccgacggg	ctcaagctcc	acctccatcg	tgggacaaa	5160
tgtggaagt	tctcatcagg	ctgaaaccta	cactgcacgg	gccaacaccc	ctgctgtata	5220
ggctaggagc	cgtccaaaat	gaggctcatc	tcacacaccc	cataactaaa	tacatcatgg	5280
catgcatgtc	ggctgacctg	gaggctgtca	ctagcacctg	gggtctggta	ggcggagctc	5340
ttgcagcttt	ggccgcatac	tgccctgacg	caggcagtg	ggctattgtg	ggcaggatca	5400
tcttgtccgg	gaagccagct	gtcgttcccg	acagggaagt	cctctaccag	gagttcgatg	5460
agatggaaga	gtgtgcctca	caacttcctt	acatcgagca	gggaatgcag	ctcgccgagc	5520
aattcaagca	aaaggcgctc	gggttgttgc	aaacggccac	caagcaagcg	gaggctgctg	5580
ctcccggtgt	ggagtccaag	tggcgagccc	ttgagacctt	ctggcggaag	cacatgtgga	5640
atthcatcag	cggaatacag	tacctagcag	gcttatccac	tctgcctgga	aaccccgcca	5700
tagcatcatt	gatggcattt	acagcttcta	tcactagccc	gtccaccacc	caaaacaccc	5760
tctgttttaa	catcttgggg	ggatgggtgg	ctgcccact	cgctcctccc	agcgctgctg	5820
cagctttcgt	ggcgccggcg	atcgccggag	cggctgttgg	cagcataggc	cttggggaagg	5880
tgctcgtgga	catcttggcg	ggctatgggg	caggggtagc	cggcgcactc	gtggccttta	5940
aggctcatgag	cggcgagggtg	ccctccaccg	aggacctggt	caacttactc	cctgccatcc	6000
tctctcctgg	tgccctggtc	gtcggggctg	tgtgcgcagc	aatactcggt	cggcacgtgg	6060
gcccgggaga	gggggctgtg	cagtggatga	accggctgat	agcgttcgct	tcgcggggta	6120
accacgtctc	ccctacgcac	tatgtgcctg	agagcgacgc	tgacgacgt	gtcactcaga	6180
tctctctag	ccttaccatc	actcaactgc	tgaagcggct	ccaccagtgg	attaatgagg	6240
actgctctac	gccatgctcc	ggctcgtggc	taagggatgt	ttgggattgg	atatgcacgg	6300
tgttgactga	cttcaagacc	tggctccagt	ccaaactcct	gccgcgggta	ccgggagctc	6360
ctttcctgtc	atgccaacgc	gggtacaagg	gagtctggcg	gggggacggc	atcatgcaaa	6420
ccacctgccc	atggcgagca	cagatcgccg	gacatgtcaa	aaacgggtcc	atgaggatcg	6480
tagggcctag	aacctgcagc	aacacgtggc	acggaacgtt	ccccatcaac	gcatacacca	6540
cgggaccttg	cacacccctc	cggcgcccca	actattccag	ggcgctatgg	cgggtggctg	6600
ctgaggagta	cgtggagggt	acgcgtgtgg	gggatttcca	ctacgtgacg	ggcatgacca	6660

-continued

ctgacaacgt aaagtgccca tgccagggtc cgccccccga attcttcacg gaggtggatg	6720
gagtgcggtt gcacagggtac gctccggcgt gcaaacctct tctacgggag gacgtcacgt	6780
tccaggtcgg gctcaaccaa tacttggtcg ggtcgcagct cccatgcgag cccgaaccgg	6840
acgtaacagt gcttacttcc atgctcaccg atccctccca cattacagca gagacggcta	6900
agcgtaggct ggctagaggg tctccccctt ctttagccag ctcacagct agccagtgtg	6960
ctgcgcttc tttgaaggcg acatgcacta cccaccatga cccccggac gctgacctca	7020
tcgaggccaa cctcttgtgg cggcaggaga tgggcggaaa catcactcgc gtggagtcat	7080
agaataaggt agtaattctg gactctttcg aaccgcttca cgcggagggg gatgagaggg	7140
agatatccgt cgcggcggag atcctgcgaa aatccaggaa gttccctca gcgttgccca	7200
tatgggcacg cccggactac aatcctccac tgctagagtc ctggaaggac cccgactacg	7260
tccctccggt ggtacacgga tgcccattgc cacctaccaa ggctcctcca ataccacctc	7320
cacggagaaa gaggacgggt gtcctgacag aatccaatgt gtcttctgcc ttggcggagc	7380
tgcacctaa gaccttcggt agctccggt cgtcggcgt tgatagcggc acggcgaccg	7440
cccttctga cctggcctcc gacgacgggt acaaaggatc cgacgttgag tcgtactcct	7500
ccatgcccc ccttgaaggg gagccggggg accccgatct cagcgacggg tcttggtcta	7560
ccgtgagtga ggaggtagt gaggatgtcg tctgctgctc aatgtcctat acgtggacag	7620
gcgcctgat cagccatgc gctgcggagg aaagtaagct gcccatcaac ccgttgagca	7680
actctttgct gcgtcaccac aacatggtct acgccacaac atccgcagc gcaagcctcc	7740
ggcagaagaa ggtcaccttt gacagattgc aagtcctgga tgatcattac cgggacgtac	7800
tcaaggagat gaaggcgaag gcgtccacag ttaaggctaa gcttctatct atagaggagg	7860
cctgcaagct gacgccccca cattcggcca aatccaaatt tggtatggg gcaaaggacg	7920
tccggaacct atccagcagg gccgttaacc acatccgctc cgtgtgggag gacttgcctg	7980
aagacactga aacaccaatt gacaccacca tcatggcaaa aagtgagggt ttctgcgtcc	8040
aaccagagaa gggaggccgc aagccagctc gccttatcgt attccagac ctgggagttc	8100
gtgtatgcga gaagatggc ctttacgacg tggctctcac ccttctcag gccgtgatgg	8160
gctctcata cggatttcaa tactccccca agcagcgggt cgagtctctg gtgaatacct	8220
ggaaatcaaa gaaatgccct atgggcttct catatgacac ccgtgtttt gactcaacgg	8280
tactgagag tgacattcgt gttgaggagt caatttacca atgttgtgac ttggcccccg	8340
agccagaca ggccataagg tcgctcacag agcggcttta catcgggggt cccctgacta	8400
actcaaaagg gcagaactgc ggttatcgcc ggtgccgcgc aagtggcgtg ctgacgacta	8460
gctgcggtaa taccctcaca tgttacttga aggccactgc agcctgtcga gctgcaaagc	8520
tccaggactg cagcatgctc gtgaacggag acgacctgt cgttatctgt gaaagcggg	8580
gaaccagga ggatgcggcg gccctacgag ccttcacgga ggctatgact aggtattccg	8640
ccccccccg ggatccgcc caaccagaat acgacctgga gctgataaca tcatgttcc	8700
ccaatgtgtc agtcgcgcac gatgcactgt gcaaaagggt atactacctc acccgtgacc	8760
ccaccacccc ccttgacagg gctgcgtggg agacagctag acacactcca atcaactctt	8820
ggctaggcaa tatcatcatg tatgcgccca ccctatgggc aaggatgatt ctgatgactc	8880
actttttctc catccttcta gctcaagagc aacttgaaaa agccctggat tgctagatct	8940
acggggcttg ctactccatt gagccacttg acctacctca gatcattgaa cgactccatg	9000

-continued

```

gtcttagcgc atttacactc cacagttact ctccagggtga gatcaatagg gtggcttcat 9060
gcctcaggaa acttggggta ccacccttgc gaacctggag acatcggggc agaagtgtcc 9120
gcgctaagct actgtcccag ggggggaggg cgcaccttg tggcagatac ctctttaact 9180
gggcagtaag gaccaagctt aaactcactc caatcccggc cgcgtcccag ctggacttgt 9240
ctggctgggt cgtcgctggt tacagcgggg gagacatata tcacagcctg tctcgtgccc 9300
gaccccgctg gtttcggtg tgcctactcc tactttctgt aggggtaggc atttacctgc 9360
tccccaacgg atgaacgggg agctaaccac tccaggcctt aagccatttc ctgttttttt 9420
tttttttttt tttttttttt tctttttttt ttttttctt ttccttcttt ttttcctttc 9480
tttttccctt ctttaatggt ggctccatct tagccctagt cacggctagc tgtgaaaggt 9540
ccgtgagccg catgactgca gagagtgctg atactggcct ctctgcagat catgt 9595

```

```

<210> SEQ ID NO 5
<211> LENGTH: 3011
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

```

```

<400> SEQUENCE: 5

```

```

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn
 1             5             10             15
Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly
 20             25             30
Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala
 35             40             45
Thr Arg Lys Ala Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro
 50             55             60
Ile Pro Lys Ala Arg Arg Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly
 65             70             75             80
Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp
 85             90             95
Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro
100            105            110
Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys
115            120            125
Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu
130            135            140
Gly Gly Ala Ala Arg Ala Leu Ala His Gly Val Arg Val Leu Glu Asp
145            150            155            160
Gly Val Asn Tyr Ala Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile
165            170            175
Phe Leu Leu Ala Leu Leu Ser Cys Leu Thr Ile Pro Ala Ser Ala Tyr
180            185            190
Glu Val Arg Asn Val Ser Gly Ile Tyr His Val Thr Asn Asp Cys Ser
195            200            205
Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Val Ile Met His Thr Pro
210            215            220
Gly Cys Val Pro Cys Val Gln Glu Gly Asn Ser Ser Arg Cys Trp Val
225            230            235            240
Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr
245            250            255
Thr Ile Arg Arg His Val Asp Leu Leu Val Gly Thr Ala Ala Phe Cys
260            265            270

```

Ser	Ala	Met	Tyr	Val	Gly	Asp	Leu	Cys	Gly	Ser	Ile	Phe	Leu	Val	Ser	
275							280				285					
Gln	Leu	Phe	Thr	Phe	Ser	Pro	Arg	Arg	His	Glu	Thr	Val	Gln	Asp	Cys	
290						295				300						
Asn	Cys	Ser	Ile	Tyr	Pro	Gly	His	Val	Ser	Gly	His	Arg	Met	Ala	Trp	
305					310					315					320	
Asp	Met	Met	Met	Asn	Trp	Ser	Pro	Thr	Thr	Ala	Leu	Val	Val	Ser	Gln	
				325					330					335		
Leu	Leu	Arg	Ile	Pro	Gln	Ala	Val	Val	Asp	Met	Val	Ala	Gly	Ala	His	
			340					345				350				
Trp	Gly	Val	Leu	Ala	Gly	Leu	Ala	Tyr	Tyr	Ser	Met	Val	Gly	Asn	Trp	
355						360						365				
Ala	Lys	Val	Leu	Ile	Val	Ala	Leu	Leu	Phe	Ala	Gly	Val	Asp	Gly	Glu	
370						375					380					
Thr	His	Thr	Thr	Gly	Arg	Val	Ala	Gly	His	Thr	Thr	Ser	Gly	Phe	Thr	
385					390					395					400	
Ser	Leu	Phe	Ser	Ser	Gly	Ala	Ser	Gln	Lys	Ile	Gln	Leu	Val	Asn	Thr	
				405					410					415		
Asn	Gly	Ser	Trp	His	Ile	Asn	Arg	Thr	Ala	Leu	Asn	Cys	Asn	Asp	Ser	
			420					425				430				
Leu	Gln	Thr	Gly	Phe	Phe	Ala	Ala	Leu	Phe	Tyr	Ala	His	Lys	Phe	Asn	
435						440					445					
Ser	Ser	Gly	Cys	Pro	Glu	Arg	Met	Ala	Ser	Cys	Arg	Pro	Ile	Asp	Trp	
450						455					460					
Phe	Ala	Gln	Gly	Trp	Gly	Pro	Ile	Thr	Tyr	Thr	Lys	Pro	Asn	Ser	Ser	
465					470					475					480	
Asp	Gln	Arg	Pro	Tyr	Cys	Trp	His	Tyr	Ala	Pro	Arg	Pro	Cys	Gly	Val	
			485						490					495		
Val	Pro	Ala	Ser	Gln	Val	Cys	Gly	Pro	Val	Tyr	Cys	Phe	Thr	Pro	Ser	
			500					505					510			
Pro	Val	Val	Val	Gly	Thr	Thr	Asp	Arg	Ser	Gly	Val	Pro	Thr	Tyr	Ser	
515							520					525				
Trp	Gly	Glu	Asn	Glu	Thr	Asp	Val	Met	Leu	Leu	Asn	Thr	Arg	Pro		
530						535				540						
Pro	Gln	Gly	Asn	Trp	Phe	Gly	Cys	Thr	Trp	Met	Asn	Ser	Thr	Gly	Phe	
545					550					555					560	
Thr	Lys	Thr	Cys	Gly	Gly	Pro	Pro	Cys	Asn	Ile	Gly	Gly	Val	Gly	Asn	
			565					570					575			
Arg	Thr	Leu	Ile	Cys	Pro	Thr	Asp	Cys	Phe	Arg	Lys	His	Pro	Glu	Ala	
			580					585				590				
Thr	Tyr	Thr	Lys	Cys	Gly	Ser	Gly	Pro	Trp	Leu	Thr	Pro	Arg	Cys	Leu	
595							600					605				
Val	Asp	Tyr	Pro	Tyr	Arg	Leu	Trp	His	Tyr	Pro	Cys	Thr	Leu	Asn	Phe	
610						615					620					
Ser	Ile	Phe	Lys	Val	Arg	Met	Tyr	Val	Gly	Gly	Val	Glu	His	Arg	Leu	
625					630					635					640	
Asn	Ala	Ala	Cys	Asn	Trp	Thr	Arg	Gly	Glu	Arg	Cys	Asn	Leu	Glu	Asp	
			645						650				655			
Arg	Asp															

-continued

690	695	700
Val Gly Ser Ala Phe 705	Val Ser Phe Ala Ile 710	Lys Trp Glu Tyr Ile Leu 715 720
Leu Leu Phe Leu Leu 725	Leu Ala Asp Ala Arg 730	Val Cys Ala Cys Leu Trp 735
Met Met Leu Leu Ile 740	Ala Gln Ala Glu Ala 745	Leu Glu Asn Leu Val 750
Val Leu Asn Ala Ala 755	Ser Val Ala Gly Ala 760	His Gly Ile Leu Ser Phe 765
Leu Val Phe Phe Cys 770	Ala Ala Trp Tyr Ile 775	Lys Gly Arg Leu Ala Pro 780
Gly Ala Ala Tyr Ala 785	Phe Tyr Gly Val Trp 790	Pro Leu Leu Leu Leu 795 800
Leu Ala Leu Pro Pro 805	Arg Ala Tyr Ala Leu 810	Asp Thr Glu Val Ala Ala 815
Ser Cys Gly Gly Val 820	Val Leu Val Gly Leu 825	Met Ala Leu Thr Leu Ser 830
Pro Tyr Tyr Lys Arg 835	Tyr Ile Ser Trp Cys 840	Met Trp Trp Leu Gln Tyr 845
Phe Leu Thr Arg Val 850	Glu Ala Gln Leu His 855	Val Trp Val Pro Pro Leu 860
Asn Val Arg Gly Gly 865	Arg Asp Ala Val Ile 870	Leu Leu Met Cys Val Val 875 880
His Pro Thr Leu Val 885	Phe Asp Ile Thr Lys 890	Leu Leu Leu Ala Ile Phe 895
Gly Pro Leu Trp Ile 900	Leu Gln Ala Ser Leu 905	Leu Lys Val Pro Tyr Phe 910
Val Arg Val Gln Gly 915	Leu Leu Arg Ile Cys 920	Ala Leu Ala Arg Lys Ile 925
Ala Gly Gly His Tyr 930	Val Gln Met Ala Ile 935	Ile Lys Leu Gly Ala Leu 940
Thr Gly Thr Tyr Val 945	Tyr Asn His Leu Thr 950	Pro Leu Arg Asp Trp Ala 955 960
His Asn Gly Leu Arg 965	Asp Leu Ala Val Ala 970	Val Glu Pro Val Val Phe 975
Ser Arg Met Glu Thr 980	Lys Leu Ile Thr Trp 985	Gly Ala Asp Thr Ala Ala 990
Cys Gly Asp Ile Ile 995	Asn Gly Leu Pro Val 1000	Ser Ala Arg Arg Gly Gln 1005
Glu Ile Leu Leu Gly 1010	Pro Ala Asp Gly Met 1015	Val Ser Lys Gly Trp Arg 1020
Leu Leu Ala Pro Ile 1025	Thr Ala Tyr Ala Gln 1030	Gln Thr Arg Gly Leu Leu 1035 1040
Gly Cys Ile Ile Thr 1045	Ser Leu Thr Gly Arg 1050	Asp Lys Asn Gln Val Glu 1055
Gly Glu Val Gln Ile 1060	Val Ser Thr Ala Thr 1065	Gln Thr Phe Leu Ala Thr 1070
Cys Ile Asn Gly Val 1075	Cys Trp Thr Val Tyr 1080	His Gly Ala Gly Thr Arg 1085
Thr Ile Ala Ser Pro 1090	Lys Gly Pro Val Ile 1095	Gln Met Tyr Thr Asn Val 1100
Asp Gln Asp Leu Val 1105	Gly Trp Pro Ala Pro 1110	Gln Gly Ser Arg Ser Leu 1115 1120

-continued

Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His
 1125 1130 1135
 Ala Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu
 1140 1145 1150
 Leu Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro
 1155 1160 1165
 Leu Leu Cys Pro Ala Gly His Ala Val Gly Leu Phe Arg Ala Ala Val
 1170 1175 1180
 Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Asn
 1185 1190 1195 1200
 Leu Gly Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro
 1205 1210 1215
 Pro Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr
 1220 1225 1230
 Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly
 1235 1240 1245
 Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe
 1250 1255 1260
 Gly Ala Tyr Met Ser Lys Ala His Gly Val Asp Pro Asn Ile Arg Thr
 1265 1270 1275 1280
 Gly Val Arg Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr
 1285 1290 1295
 Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile
 1300 1305 1310
 Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly
 1315 1320 1325
 Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val
 1330 1335 1340
 Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Ser His Pro
 1345 1350 1355 1360
 Asn Ile Glu Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr
 1365 1370 1375
 Gly Lys Ala Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile
 1380 1385 1390
 Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val
 1395 1400 1405
 Ala Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser
 1410 1415 1420
 Val Ile Pro Thr Ser Gly Asp Val Val Val Val Ser Thr Asp Ala Leu
 1425 1430 1435 1440
 Met Thr Gly Phe Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr
 1445 1450 1455
 Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile
 1460 1465 1470
 Glu Thr Thr Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg
 1475 1480 1485
 Gly Arg Thr Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro
 1490 1495 1500
 Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys
 1505 1510 1515 1520
 Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr
 1525 1530 1535

-continued

Val Arg Leu Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln			
1540	1545	1550	
Asp His Leu Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile			
1555	1560	1565	
Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Phe Pro			
1570	1575	1580	
Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro			
1585	1590	1595	1600
Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro			
1605	1610	1615	
Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln			
1620	1625	1630	
Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Thr Cys			
1635	1640	1645	
Met Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly			
1650	1655	1660	
Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val			
1665	1670	1675	1680
Val Ile Val Gly Arg Ile Val Leu Ser Gly Lys Pro Ala Ile Ile Pro			
1685	1690	1695	
Asp Arg Glu Val Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ser			
1700	1705	1710	
Gln His Leu Pro Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe			
1715	1720	1725	
Lys Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Ser Arg His Ala Glu			
1730	1735	1740	
Val Ile Thr Pro Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Val Phe			
1745	1750	1755	1760
Trp Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala			
1765	1770	1775	
Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala			
1780	1785	1790	
Phe Thr Ala Ala Val Thr Ser Pro Leu Thr Thr Gly Gln Thr Leu Leu			
1795	1800	1805	
Phe Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro Gly			
1810	1815	1820	
Ala Ala Thr Ala Phe Val Gly Ala Gly Leu Ala Gly Ala Ala Ile Gly			
1825	1830	1835	1840
Ser Val Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly			
1845	1850	1855	
Ala Gly Val Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu			
1860	1865	1870	
Val Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser			
1875	1880	1885	
Pro Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg			
1890	1895	1900	
His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile			
1905	1910	1915	1920
Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro			
1925	1930	1935	
Glu Ser Asp Ala Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu Thr			
1940	1945	1950	
Val Thr Gln Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys			

-continued

1955					1960					1965					
Thr	Thr	Pro	Cys	Ser	Gly	Ser	Trp	Leu	Arg	Asp	Ile	Trp	Asp	Trp	Ile
1970						1975					1980				
Cys	Glu	Val	Leu	Ser	Asp	Phe	Lys	Thr	Trp	Leu	Lys	Ala	Lys	Leu	Met
1985				1990						1995				2000	
Pro	Gln	Leu	Pro	Gly	Ile	Pro	Phe	Val	Ser	Cys	Gln	Arg	Gly	Tyr	Arg
			2005						2010					2015	
Gly	Val	Trp	Arg	Gly	Asp	Gly	Ile	Met	His	Thr	Arg	Cys	His	Cys	Gly
		2020						2025					2030		
Ala	Glu	Ile	Thr	Gly	His	Val	Lys	Asn	Gly	Thr	Met	Arg	Ile	Val	Gly
	2035					2040						2045			
Pro	Arg	Thr	Cys	Arg	Asn	Met	Trp	Ser	Gly	Thr	Phe	Pro	Ile	Asn	Ala
	2050					2055					2060				
Tyr	Thr	Thr	Gly	Pro	Cys	Thr	Pro	Leu	Pro	Ala	Pro	Asn	Tyr	Lys	Phe
2065				2070						2075				2080	
Ala	Leu	Trp	Arg	Val	Ser	Ala	Glu	Glu	Tyr	Val	Glu	Ile	Arg	Arg	Val
		2085						2090						2095	
Gly	Asp	Phe	His	Tyr	Val	Ser	Gly	Met	Thr	Thr	Asp	Asn	Leu	Lys	Cys
		2100						2105					2110		
Pro	Cys	Gln	Ile	Pro	Ser	Pro	Glu	Phe	Phe	Thr	Glu	Leu	Asp	Gly	Val
	2115						2120						2125		
Arg	Leu	His	Arg	Phe	Ala	Pro	Pro	Cys	Lys	Pro	Leu	Leu	Arg	Glu	Glu
	2130					2135					2140				
Val	Ser	Phe	Arg	Val	Gly	Leu	His	Glu	Tyr	Pro	Val	Gly	Ser	Gln	Leu
2145				2150						2155				2160	
Pro	Cys	Glu	Pro	Glu	Pro	Asp	Val	Ala	Val	Leu	Thr	Ser	Met	Leu	Thr
		2165						2170						2175	
Asp	Pro	Ser	His	Ile	Thr	Ala	Glu	Ala	Ala	Gly	Arg	Arg	Leu	Ala	Arg
		2180						2185					2190		
Gly	Ser	Pro	Pro	Ser	Met	Ala	Ser	Ser	Ser	Ala	Ser	Gln	Leu	Ser	Ala
	2195						2200						2205		
Pro	Ser	Leu	Lys	Ala	Thr	Cys	Thr	Ala	Asn	His	Asp	Ser	Pro	Asp	Ala
	2210					2215						2220			
Glu	Leu	Ile	Glu	Ala	Asn	Leu	Leu	Trp	Arg	Gln	Glu	Met	Gly	Gly	Asn
2225				2230						2235				2240	
Ile	Thr	Arg	Val	Glu	Ser	Glu	Asn	Lys	Val	Val	Ile	Leu	Asp	Ser	Phe
		2245						2250						2255	
Asp	Pro	Leu	Val	Ala	Glu	Glu	Asp	Glu	Arg	Glu	Val	Ser	Val	Pro	Ala
		2260						2265					2270		
Glu	Ile	Leu	Arg	Lys	Ser	Arg	Arg	Phe	Ala	Arg	Ala	Leu	Pro	Val	Trp
	2275						2280						2285		
Ala	Arg	Pro	Asp	Tyr	Asn	Pro	Pro	Leu	Val	Glu	Thr	Trp	Lys	Lys	Pro
	2290					2295						2300			
Asp	Tyr	Glu	Pro	Pro	Val	Val	His	Gly	Cys	Pro	Leu	Pro	Pro	Pro	Arg
2305				2310						2315				2320	
Ser	Pro	Pro	Val	Pro	Pro	Pro	Arg	Lys	Lys	Arg	Thr	Val	Val	Leu	Thr
		2325							2330					2335	
Glu	Ser	Thr	Leu	Ser	Thr	Ala	Leu	Ala	Glu	Leu	Ala	Thr	Lys	Ser	Phe
		2340						2345					2350		
Gly	Ser	Ser	Ser	Thr	Ser	Gly	Ile	Thr	Gly	Asp	Asn	Thr	Thr	Thr	Ser
	2355						2360						2365		
Ser	Glu	Pro	Ala	Pro	Ser	Gly	Cys	Pro	Pro	Asp	Ser	Asp	Val	Glu	Ser
	2370					2375					2380				

-continued

Tyr Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu
 2385 2390 2395 2400
 Ser Asp Gly Ser Trp Ser Thr Val Ser Ser Gly Ala Asp Thr Glu Asp
 2405 2410 2415
 Val Val Cys Cys Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr
 2420 2425 2430
 Pro Cys Ala Ala Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn
 2435 2440 2445
 Ser Leu Leu Arg His His Asn Leu Val Tyr Ser Thr Thr Ser Arg Ser
 2450 2455 2460
 Ala Cys Gln Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu
 2465 2470 2475 2480
 Asp Ser His Tyr Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ala Ser
 2485 2490 2495
 Lys Val Lys Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Thr
 2500 2505 2510
 Pro Pro His Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val
 2515 2520 2525
 Arg Cys His Ala Arg Lys Ala Val Ala His Ile Asn Ser Val Trp Lys
 2530 2535 2540
 Asp Leu Leu Glu Asp Ser Val Thr Pro Ile Asp Thr Thr Ile Met Ala
 2545 2550 2555 2560
 Lys Asn Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro
 2565 2570 2575
 Ala Arg Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys
 2580 2585 2590
 Met Ala Leu Tyr Asp Val Val Ser Lys Leu Pro Leu Ala Val Met Gly
 2595 2600 2605
 Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu
 2610 2615 2620
 Val Gln Ala Trp Lys Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp
 2625 2630 2635 2640
 Thr Arg Cys Phe Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr Glu
 2645 2650 2655
 Glu Ala Ile Tyr Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val Ala
 2660 2665 2670
 Ile Lys Ser Leu Thr Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn
 2675 2680 2685
 Ser Arg Gly Glu Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val
 2690 2695 2700
 Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg
 2705 2710 2715 2720
 Ala Ala Cys Arg Ala Ala Gly Leu Gln Asp Cys Thr Met Leu Val Cys
 2725 2730 2735
 Gly Asp Asp Leu Val Val Ile Cys Glu Ser Ala Gly Val Gln Glu Asp
 2740 2745 2750
 Ala Ala Ser Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala
 2755 2760 2765
 Pro Pro Gly Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr
 2770 2775 2780
 Ser Cys Ser Ser Asn Val Ser Val Ala His Asp Gly Ala Gly Lys Arg
 2785 2790 2795 2800

-continued

tcccagcttc	cgcttatgaa	gtgcgcaacg	tgtccgggat	ataccatgtc	acgaacgact	960
gctccaaactc	aagcattgtg	tatgaggcag	cggacgtgat	catgcatact	cccgggtgcg	1020
tgccctgtgt	tcaggagggg	aacagctccc	gttgcctggg	agcgcctcact	cccacgctcg	1080
cggccaggaa	tgccagcgtc	cccactacga	caatacgacg	ccacgtcgac	ttgctcgttg	1140
ggacggctgc	tttctgctcc	gctatgtacg	tgggggatct	ctgcggatct	attttcctcg	1200
tctcccagct	gttcaccttc	tcgcctcgcc	ggcatgagac	agtgcaggac	tgcaactgct	1260
caatctatcc	cggccatgta	tcaggtcacc	gcattggcttg	ggatatgatg	atgaactggg	1320
cacctacaac	agccctagtgt	gtgtcgcagt	tgtcccggtg	cccacaagct	gtcgtggaca	1380
tggtagcggg	ggcccactgg	ggagtcctgg	cgggccttgc	ctactattcc	atggtaggga	1440
actgggctaa	ggttctgatt	gtggcgctac	tctttgccgg	cgttgacggg	gagaccacaca	1500
cgcacggggag	ggtggccggc	cacaccacct	cggggttcac	gtcccttttc	tcatctgggg	1560
cgtctcagaa	aatccagctt	gtgaatacca	acggcagctg	gcacatcaac	aggactgccc	1620
taaatgcaa	tgactccctc	caaactgggt	tctttgccgc	gctgttttac	gcacacaagt	1680
tcaactcgtc	cgggtgcccg	gagcgcatgg	ccagctgccg	ccccattgac	tggttcgccc	1740
aggggtgggg	ccccatcacc	tatactaagc	ctaacagctc	ggatcagagg	ccttattgct	1800
ggcattacgc	gcctcgaccg	tgtggtgtcg	taccgcgctc	gcagggtgtg	ggtccagtgt	1860
attgtttcac	cccaagccct	gttgtggtgg	ggaccaccga	tcgttccggg	gtccctacgt	1920
atagctgggg	ggagaatgag	acagacgtga	tgctcctcaa	caacacgcgt	ccgccacaag	1980
gcaactggtt	cggctgtaca	tggatgaata	gtactggggt	cactaagacg	tcgggagggtc	2040
ccccgtgtaa	catcgggggg	gtcggtaacc	gcaccttgat	ctgccccacg	gactgtcttc	2100
ggaagcacc	cgcagctact	tacacaaaat	gtggctcggg	gccttggttg	acacctaggt	2160
gcctagtaga	ctaccatatac	aggctttggc	actaccctcg	cactctcaat	ttttccatct	2220
ttaagggttag	gatgtatgtg	gggggcgtgg	agcacaggct	caatgccgca	tgcaattgga	2280
ctcgaggaga	gcgctgtaac	ttggaggaca	gggatagggtc	agaactcagc	ccgctgctgc	2340
tgtctacaac	agagtggcag	atactgcctc	gtgctttcac	cacctaccg	gctttatcca	2400
ctggtttgat	ccatctccat	cagaacatcg	tggacgtgca	atacctgtac	ggtgtagggt	2460
cagcgtttgt	ctcctttgca	atcaaatggg	agtaacatct	gttgcttttc	cttctcctgg	2520
cagacgcgcg	cgtgtgtgcc	tgcttgtgga	tgatgctgct	gatagcccag	gctgaggccg	2580
ccttagagaa	cttgggtggtc	ctcaatgcgg	cgtccgtggc	cggagcgcac	ggtattctct	2640
cctttcttgt	gttctctgc	gcgcctggt	acattaaggg	caggctggct	cctggggcgg	2700
cgtatgcttt	ttatggcgta	tggccgctgc	tcctgctcct	actggcggtta	ccaccacgag	2760
catatgcact	ggacacggag	gtggccgcgt	cgtgtggcgg	cgtgtgtctt	gtcgggttaa	2820
tggcgctgac	tctgtcgcca	tattacaagc	gctatatcag	ctggtgcatg	tggtagcttc	2880
agtattttct	gaccagagta	gaagcgcaac	tgcacgtgtg	ggttcccccc	ctcaacgtcc	2940
ggggggggcg	cgatgccgtc	atcttactca	tgtgtgtagt	acaccgcacc	ctggtatttg	3000
acatcaccaa	actactcctg	gccatcttcg	gacccctttg	gattcttcaa	gccagtttgc	3060
ttaaagtccc	ctacttcgtg	cgcgttcaag	gccttctcct	gatctgcgcg	ctagcgcgga	3120
agatagccgg	aggtcattac	gtgcaaatgg	ccatcatcaa	gttagggcg	cttactggca	3180
cctatgtgta	taaccatctc	acccctcttc	gagactgggc	gcacaacggc	ctgcgagatc	3240

-continued

tggcgtggc	tgtggaacca	gtcgtcttct	cccgaatgga	gaccaagctc	atcacgtggg	3300
gggcagatac	cgccgcgtgc	ggtgacatca	tcaacggctt	gcccgtctct	gcccgtaggg	3360
gccaggagat	actgcttggg	ccagccgacg	gaatgggtctc	caaggggtgg	aggttgctgg	3420
cggccatcac	ggcgtacgcc	cagcagacga	gaggcctcct	aggggtgtata	atcaccagcc	3480
tgactggccg	ggacaaaaac	caagtggagg	gtgagggtcca	gacgtgtca	actgctaccc	3540
aaaccttcct	ggcaactgtc	atcaatgggg	tatgctggac	tgtctaccac	ggggccggaa	3600
cgaggaccat	cgcacacccc	aaggggtcctg	tcacccagat	gtataccaat	gtggaccaag	3660
accttgctgg	ctggcccgtc	cctcaaggtt	cccgtcatt	gacaccctgt	acctgctggc	3720
cctcgagacct	ttacctggtc	acgaggcacg	ccgatgtcat	tcccgtgcgc	cggcgaggtg	3780
atagcagggg	tagctgtctt	tcgccccggc	ccatttccta	cttgaaaggc	tcctcggggg	3840
gtccgtgtt	gtgccccgcg	ggacacgcgc	tgggcctatt	cagggccgcg	gtgtgcaccc	3900
gtggagtggc	taaagcgggtg	gactttatcc	ctgtggagaa	cctagggaca	accatgagat	3960
ccccggtgtt	cacggacaac	tcctctccac	cagcagtgcc	ccagagcttc	caggtggccc	4020
acctgcacgc	tcccaccggc	agcggtaaga	gcaccaaggt	cccggctgcg	tacgcagccc	4080
aggggtacaa	ggtgttgggtg	ctcaaccctc	ctgttgctgc	aacgtgggc	tttgggtgctt	4140
acatgtccaa	ggcccatggg	gttgatccta	atatcaggac	cgggggtgaga	acaattacca	4200
ctggcagccc	catcacgtac	tcacctacg	gcaagtctct	tgcgcacggc	gggtgctcag	4260
gaggtgctta	tgacataata	atttgtagcg	agtgcactc	cacggatgcc	acatccatct	4320
tgggcatcgg	cactgtcctt	gaccaagcag	agactgcggg	ggcgagactg	gttggtgctc	4380
ccactgtac	ccctccgggc	tcctgactg	tgtcccatcc	taacatcgag	gaggttgctc	4440
tgtccaccac	cggagagatc	cccttttacg	gcaaggctat	ccccctcgag	gtgatcaagg	4500
ggggaagaca	tctcatcttc	tgccactcaa	agaagaagtg	cgacgagctc	gccgcgaagc	4560
tgggtgcatt	gggcatcaat	gccgtggcct	actaccgcgg	tcttgacgtg	tctgtcatcc	4620
cgaccagcgg	cgatgttgct	gtcgtgtcga	ccgatgctct	catgactggc	tttaccggcg	4680
acttcgactc	tgtgatagac	tgcaaacagt	gtgtcactca	gacagtcat	ttcagccttg	4740
accttacctt	taccattgag	acaaccacgc	tccccagga	tgctgtctcc	aggactcaac	4800
gccggggcag	gactggcagg	gggaagccag	gcacttatag	atttggtggc	ccgggggagc	4860
gcccctccgg	catgttcgac	tcgtccgtcc	tctgtgagtg	ctatgacgcg	ggctgtgctt	4920
gggatgagct	cacgcccgcg	gagactacag	ttaggctacg	agcgtacatg	aacaccccgg	4980
ggcttcccgt	gtgccaggac	catcttgaat	tttgggaggg	cgtctttacg	ggcctcactc	5040
atatagatgc	ccacttttta	tcccagacaa	agcagagtgg	ggagaacttt	ccttacctgg	5100
tagcgtacca	agccaccgtg	tcgctaggg	ctcaagcccc	tccccatcg	tgggaccaga	5160
tgtggaagtg	tttgatccgc	cttaaaccca	ccctccatgg	gccaacaccc	ctgctataca	5220
gactggggcg	tgttcagaat	gaagtcaccc	tgacgcaccc	aatcaccaaa	tacatcatga	5280
catgcatgtc	ggcgcacctg	gaggtcgtca	cgagcacctg	gggtgctcgt	ggcggcgtcc	5340
tggctgctct	ggccgcgtat	tgccctgtcaa	caggctgcgt	ggctcatagt	ggcaggatcg	5400
tcttgctccg	gaagccggca	attataacctg	acagggaggt	tctctaccag	gagttcgatg	5460
agatggaaga	gtgctctcag	cacttacctg	acatcgagca	agggatgatg	ctcgtgagc	5520
agttcaagca	gaaggccctc	ggcctcctgc	agaccgcgtc	ccgccatgca	gaggttatca	5580
ccccgtctgt	ccagaccaac	tggcagaaac	tcgaggtctt	tggggcgaag	cacatgtgga	5640

-continued

atttcatcag	tgggatacaa	tacttgccgg	gcctgtcaac	gctgcctggt	aaccccgcca	5700
ttgcttcatt	gatggetttt	acagctgccg	tcaccagccc	actaaccact	ggccaaaccc	5760
tcctcttcaa	catattgggg	gggtgggtgg	ctgcccagct	cgccgcccc	ggtgccgcta	5820
ctgcctttgt	gggtgctggc	ctagctggcg	cgcctcatcg	cagcgttgga	ctggggaagg	5880
tcctcgtgga	cattcttgca	gggtatggcg	cgggcgtggc	gggagctctt	gtagcattca	5940
agatcatgag	cggtagggtc	ccctccacgg	aggacctggt	caatctgctg	cccgccatcc	6000
tctcgctcgg	agcccttgta	gtcggtgtgg	tctgcgcagc	aatactgcgc	cggcacgttg	6060
gccccggcga	gggggcagtg	caatggatga	accggctaata	agccttcgcc	tccccgggga	6120
accatgtttc	ccccacgcac	tacgtgccgg	agagcgatgc	agccgcccgc	gtcactgcca	6180
tactcagcag	cctcactgta	accagctccc	tgaggcgact	gcatcagtgg	ataagctcgg	6240
agtgtaccac	tccatgctcc	ggttcctggc	taaggacat	ctgggactgg	atatgcgagg	6300
tgctgagcga	ctttaagacc	tggctgaaag	ccaagctcat	gccacaactg	cctgggattc	6360
cctttgtgtc	ctgccagcgc	gggtataggg	gggtctggcg	aggagacggc	attatgcaca	6420
ctcgtgccca	ctgtggagct	gagatcactg	gacatgtcaa	aaacgggacg	atgaggatcg	6480
tcggtcctag	gacctgcagg	aacatgtgga	gtgggacggt	ccccattaac	gcctacacca	6540
cgggcccctg	tactccccct	cctgcgcgga	actataagtt	cgcgctgtgg	aggggtgtctg	6600
cagaggaata	cgtggagata	aggcgggtgg	gggacttcca	ctacgtatcg	ggtatgacta	6660
ctgacaatct	taaatgcccc	tgccagatcc	catcgcccga	atttttcaca	gaattggacg	6720
gggtgcgcct	acacaggttt	gcgccccctt	gcaagccctt	gctgcgggag	gagggtatcat	6780
tcagagtagg	actccacgag	taccgggtgg	ggtcgcaatt	accttgcgag	cccgaaccgg	6840
acgtagccgt	gttgacgtcc	atgctcactg	atccctccca	tataacagca	gaggcggccg	6900
ggagaagggt	ggcgagaggg	tcacccccct	ctatggccag	ctcctcggct	agccagctgt	6960
ccgctccatc	tctcaaggca	acttgccacc	ccaacatga	ctccctgac	gccgagctca	7020
tagaggctaa	cctcctgtgg	aggcaggaga	tgggcggcaa	catcaccagg	ggtgagtcag	7080
agaacaaagt	ggtgattctg	gactccttcg	atccgcttgt	ggcagaggag	gatgagcggg	7140
aggtctccgt	acctgcagaa	attctgcgga	agtctcgag	attcgcccgg	gccctgcccg	7200
tctgggcgcg	gccggactac	aacccccccg	tagtagagac	gtggaaaaag	cctgactacg	7260
aaccacctgt	ggtccatggc	tgcccgttac	cacctccacg	gtccctcctt	gtgcctccgc	7320
ctcgaaaaaa	gcgtacgggt	gtcctcaccg	aatcaaccct	atctactgcc	ttggccgagc	7380
ttgccaccaa	aagtttttgc	agctcctcaa	cttcggcat	tacgggcgac	aatacgacaa	7440
catcctctga	gcccgcctct	tctggctgcc	cccccgactc	cgcggttgag	tctattctct	7500
ccatgcccc	cctggagggg	gagcctgggg	atccggatct	cagcgacggg	tcatggtcga	7560
cggtcagtag	tggggccgac	acggaagatg	tcgtgtgctg	ctcaatgtct	tattcctgga	7620
caggcgcact	cgtcaccccc	tgcgctgcgg	aagaacaaaa	actgcccata	aacgcactga	7680
gcaactcgtt	gtacagccat	cacaatctgg	tgtattccac	cacttcacgc	agtgtttgcc	7740
aaaggcagaa	gaaagtcaca	tttgacagac	tgcaagttct	ggacagccat	taccaggacg	7800
tgctcaagga	ggtcaaagca	gcggcgtcaa	aagtgaaggc	taacttgcta	tccgtagagg	7860
aagcttgca	cctgacgccc	ccacattcag	ccaaatccaa	gtttggctat	ggggcaaaag	7920
acgtccgttg	ccatgccaga	aaggccgtag	cccacatcaa	ctccgtgtgg	aaagaccttc	7980

-continued

tggaagacag tgtaacacca atagacacta ccatcatggc caagaacgag gttttctgcg	8040
ttcagcctga gaaggggggt cgtaagccag ctctgtctcat cgtgttcccc gacctgggcg	8100
tgcgcgtgtg cgagaagatg gccctgtacg acgtgggttag caagctcccc ctggccgtga	8160
tgggaagctc ctacggattc caatactcac caggacagcg ggttgaattc ctctgtgcaag	8220
cgtggaagtc caagaagacc ccgatgggggt tctcgtatga taccgctgt tttgactcca	8280
cagtcactga gagcgacatc cgtacggagg aggcaattta ccaatgttgt gacctggacc	8340
cccaagcccc cgtggccatc aagtccctca ctgagaggct ttatgttggg ggccctctta	8400
ccaattcaag gggggaaaac tgcggtacc gcaggtgccg cgcgagcggc gtactgacaa	8460
ctagctgtgg taacaccctc acttgctaca tcaaggcccc ggagcctgt cgagccgag	8520
ggctccagga ctgcaccatg ctctgtgtgt gcgacgactt agtcgttacc tgtgaaagt	8580
cgggggtcca ggaggacgcg gcgagcctga gagccttcac ggaggctatg accaggtact	8640
ccgccccccc cggggacccc ccacaaccag aatacgactt ggagcttata acatcatgct	8700
cctccaacgt gtcagtcgcc cacgacggcg ctggaagag ggtctactac cttaccctg	8760
accctacaac cccctcgcg agagccgctt gggagacagc aagacacact ccagtcaatt	8820
cctggctagg caacataatc atgtttgccc ccacactgtg ggcgaggatg atactgatga	8880
cccatttctt tagcgtctc atagccaggg atcagcttga acaggctctt aactgtgaga	8940
tctacggagc ctgctactcc atagaaccac tggatctacc tccaatcatt caaagactcc	9000
atggcctcag cgcattttca ctccacagtt actctccagg tgaatcaat aggggtggccg	9060
catgcctcag aaaacttggg gtcccgcctt tgcgagcttg gagacaccgg gcccgagcg	9120
tccgcgctag gcttctgtcc agaggaggca gggctgctat atgtggcaag tacctcttca	9180
actgggcagt aagaacaaag ctcaaaactca ctccaatagc ggcgctggc cggtggact	9240
tgtccggttg gttcacggct ggctacagcg ggggagacat ttatcacagc gtgtctcatg	9300
cccggccccg ctggttctgg ttttgctac tctgtctgc tgcaggggta ggcactacc	9360
tctccccc aa cgatgaagg ttggggtaaa cactccggcc tcttaagcca tttctgttt	9420
tttttttttt tttttttttt tttttctttt tttttttctt tcttttctt ctttttttcc	9480
ttttttttt cttcttttaa tgggtggctc atcttagccc tagtcacggc tagctgtgaa	9540
aggctccgtga gcgcgatgac tgcagagagt gctgatactg gcctctctgc agatcatgt	9599

<210> SEQ ID NO 7
 <211> LENGTH: 31
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 7

ggctacagcg gggggagaca tttatcacag c	31
------------------------------------	----

<210> SEQ ID NO 8
 <211> LENGTH: 30
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 8

tcattgaggct caggacctt tcacagctag	30
----------------------------------	----

<210> SEQ ID NO 9
 <211> LENGTH: 38
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

-continued

<400> SEQUENCE: 9

gtccaagctt atcacagcgt gtctcatgcc cggccccg 38

<210> SEQ ID NO 10

<211> LENGTH: 39

<212> TYPE: DNA

<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 10

cgtctctaga ggacctttca cagctagccg tgactaggg 39

<210> SEQ ID NO 11

<211> LENGTH: 38

<212> TYPE: DNA

<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 11

tgaaggttgg ggtaaacact ccggcctctt aggccatt 38

<210> SEQ ID NO 12

<211> LENGTH: 35

<212> TYPE: DNA

<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 12

acatgatctg cagagaggcc agtatcagca ctctc 35

<210> SEQ ID NO 13

<211> LENGTH: 46

<212> TYPE: DNA

<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 13

gtccaagctt acgcgtaaac actccggcct ccttaagcca ttcttg 46

<210> SEQ ID NO 14

<211> LENGTH: 47

<212> TYPE: DNA

<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 14

cgtctctaga catgatctgc agagaggcca gtatcagcac tctctgc 47

<210> SEQ ID NO 15

<211> LENGTH: 67

<212> TYPE: DNA

<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 15

ttttttttgc ggccgcta atagccagcc ccctgatggg ggcgacactc 60

caccatg 67

<210> SEQ ID NO 16

<211> LENGTH: 30

<212> TYPE: DNA

<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 16

actgtcttca cgcagaaagc gtctagccat 30

<210> SEQ ID NO 17

-continued

```

<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 17
cgtctctaga caggaaatgg cttaagaggc cggagtgttt acc          43

<210> SEQ ID NO 18
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 18
gcctattggc ctggagtggg tagctc          26

<210> SEQ ID NO 19
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 19
aggatggcct taaggcctgg agtgggttagc tccccgttca          40

<210> SEQ ID NO 20
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 20
cgtcatcgat cctcagcggg catatgcact ggacacgga          39

<210> SEQ ID NO 21
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 21
catgcaccag ctgatatagc gcttgtaata tg          32

<210> SEQ ID NO 22
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 22
tccgtagagg aagcttgagc cctgacgccc          30

<210> SEQ ID NO 23
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 23
gtacttgcca catatagcag ccctgctccc tctg          34

<210> SEQ ID NO 24
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 24
cagaggaggc agggctgcta tatgtggcaa gtac          34

```

-continued

<210> SEQ ID NO 25
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 25

cgtctctaga caggaaatgg cttaagaggc cggagtgttt acc 43

<210> SEQ ID NO 26
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 26

tgcaattgga ctcgaggaga gcgctgtaac ttggag 36

<210> SEQ ID NO 27
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 27

cgggtccaagg catatgctcg tgggtggaac gccag 35

<210> SEQ ID NO 28
<211> LENGTH: 35
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 28

Ala Gly Val Asp Gly Glu Thr His Thr Thr Gly Arg Val Ala Gly His
1 5 10 15

Thr Thr Ser Gly Phe Thr Ser Leu Phe Ser Ser Gly Ala Ser Gln Lys
20 25 30

Ile Gln Leu
35

<210> SEQ ID NO 29
<211> LENGTH: 19
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 29

Gly Trp Gly Pro Ile Thr Tyr Thr Lys Pro Asn Ser Ser Asp Gln Arg
1 5 10 15

Pro Tyr Cys

<210> SEQ ID NO 30
<211> LENGTH: 35
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 30

Ala Gly Val Asp Gly Glu Thr His Thr Thr Gly Arg Val Ala Gly His
1 5 10 15

Thr Thr Ser Arg Phe Thr Ser Leu Phe Ser Ser Gly Ala Ser Gln Lys
20 25 30

Ile Gln Leu
35

<210> SEQ ID NO 31
<211> LENGTH: 19

-continued

```

<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 31

Gly Trp Gly Pro Ile Thr Tyr Thr Glu Pro Asn Ser Ser Asp Gln Arg
  1             5             10             15

Pro Tyr Cys

<210> SEQ ID NO 32
<211> LENGTH: 35
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 32

Ala Gly Val Asp Gly Glu Thr His Thr Thr Gly Arg Val Val Gly His
  1             5             10             15

Thr Thr Ser Gly Phe Thr Ser Leu Phe Ser Ser Gly Ala Ser Gln Lys
             20             25             30

Ile Gln Leu
             35

<210> SEQ ID NO 33
<211> LENGTH: 19
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 33

Gly Trp Gly Pro Ile Thr Tyr Thr Gly Pro Asn Ser Ser Asp Gln Arg
  1             5             10             15

Pro Tyr Cys

<210> SEQ ID NO 34
<211> LENGTH: 35
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 34

Ala Gly Val Asp Gly Glu Thr His Thr Thr Gly Arg Val Val Gly Arg
  1             5             10             15

Thr Thr Ser Gly Phe Thr Ser Leu Phe Ser Ser Gly Ala Ser Gln Lys
             20             25             30

Ile Gln Leu
             35

<210> SEQ ID NO 35
<211> LENGTH: 19
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 35

Gly Trp Gly Pro Ile Ala Tyr Thr Glu Pro Asn Ser Ser Asp Gln Arg
  1             5             10             15

Pro Tyr Cys

<210> SEQ ID NO 36
<211> LENGTH: 35
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 36

Ala Gly Val Asp Gly Thr Thr Tyr Thr Ser Gly Gly Val Ala Gly Arg

```

```

1           5           10          15
Thr Thr Ser Gly Phe Thr Ser Leu Phe Ser Pro Gly Ala Ser Gln Lys
      20           25           30

Ile Gln Leu
      35

<210> SEQ ID NO 37
<211> LENGTH: 35
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 37

Thr Gly Val Asp Gly Thr Thr Tyr Thr Ser Gly Gly Ala Ala Gly Arg
  1           5           10           15

Thr Thr Ser Gly Phe Thr Ser Leu Phe Ser Ser Gly Ala Ser Gln Lys
      20           25           30

Ile Gln Leu
      35

<210> SEQ ID NO 38
<211> LENGTH: 35
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 38

Thr Gly Val Asp Gly Thr Thr Tyr Thr Ser Gly Gly Val Ala Gly Arg
  1           5           10           15

Thr Thr Ser Gly Phe Thr Ser Leu Phe Ser Ser Gly Ala Ser Gln Lys
      20           25           30

Ile Gln Leu
      35

<210> SEQ ID NO 39
<211> LENGTH: 19
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 39

Gly Trp Gly Pro Ile Thr His Thr Glu Pro Asn Ser Ser Asp Gln Arg
  1           5           10           15

Pro Tyr Cys

<210> SEQ ID NO 40
<211> LENGTH: 19
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 40

Gly Trp Gly Pro Ile Thr Tyr Thr Gly Pro Asp Ser Leu Asp Gln Arg
  1           5           10           15

Pro Tyr Cys

<210> SEQ ID NO 41
<211> LENGTH: 35
<212> TYPE: PRT
<213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 41

Ala Gly Val Asp Gly Ala Thr Tyr Thr Ser Gly Gly Val Ala Gly Arg
  1           5           10           15

```

-continued

Thr Thr Ser Gly Phe Thr Ser Leu Phe Ser Ser Gly Ala Ser Gln Lys
 20 25 30

Ile Gln Leu
 35

<210> SEQ ID NO 42
 <211> LENGTH: 19
 <212> TYPE: PRT
 <213> ORGANISM: Hepatitis C virus
 <400> SEQUENCE: 42

Gly Trp Gly Pro Ile Thr Tyr Thr Glu Pro Asn Ser Pro Asp Gln Arg
 1 5 10 15

Pro Tyr Cys

<210> SEQ ID NO 43
 <211> LENGTH: 35
 <212> TYPE: PRT
 <213> ORGANISM: Hepatitis C virus
 <400> SEQUENCE: 43

Ala Gly Val Asp Gly Lys Thr Tyr Thr Ser Gly Gly Ala Ala Ser His
 1 5 10 15

Thr Thr Ser Arg Phe Thr Ser Leu Phe Ser Pro Gly Ala Ser Gln Arg
 20 25 30

Ile Gln Leu
 35

<210> SEQ ID NO 44
 <211> LENGTH: 19
 <212> TYPE: PRT
 <213> ORGANISM: Hepatitis C virus
 <400> SEQUENCE: 44

Gly Trp Gly Pro Ile Thr Tyr Thr Glu Ser Gly Ser Arg Asp Gln Arg
 1 5 10 15

Pro Tyr Cys

<210> SEQ ID NO 45
 <211> LENGTH: 35
 <212> TYPE: PRT
 <213> ORGANISM: Hepatitis C virus
 <400> SEQUENCE: 45

Ala Gly Val Asp Gly Glu Thr Tyr Thr Ser Gly Gly Ala Ala Ser His
 1 5 10 15

Thr Thr Ser Thr Leu Ala Ser Leu Phe Ser Pro Gly Ala Ser Gln Arg
 20 25 30

Ile Gln Leu
 35

<210> SEQ ID NO 46
 <211> LENGTH: 19
 <212> TYPE: PRT
 <213> ORGANISM: Hepatitis C virus
 <400> SEQUENCE: 46

Gly Trp Gly Pro Ile Thr Tyr Thr Glu Pro Asp Ser Pro Asp Gln Arg
 1 5 10 15

Pro Tyr Cys

-continued

<210> SEQ ID NO 47
 <211> LENGTH: 341
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 47

```
gccagcccc gattgggggc gacactccac catagatcac tcccctgtga ggaactactg    60
tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgcgtgcag cctccaggac    120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag    180
gacgaccggg tcctttcttg gatcaaccg ctcaatgcct ggagatttgg gcgtgcccc    240
gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtgttact gcctgatagg    300
gtgcttgcca gtgccccggg aggtctcgta gaccgtgcac c                        341
```

<210> SEQ ID NO 48
 <211> LENGTH: 341
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 48

```
gccagcccc tgatgggggc gacactccac catgaatcac tcccctgtga ggaactactg    60
tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgcgtgcag cctccaggac    120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag    180
gacgaccggg tcctttcttg gatcaaccg ctcaatgcct ggagatttgg gcgtgcccc    240
gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtgttact gcctgatagg    300
gtgcttgcca gtgccccggg aggtctcgta gaccgtgcac c                        341
```

<210> SEQ ID NO 49
 <211> LENGTH: 341
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 49

```
gccagcccc tgatgggggc gacactccac catgaatcac tcccctgtga ggaactactg    60
tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgcgtgcag cctccaggac    120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag    180
gacgaccggg tcctttcttg gataaaccg ctcaatgcct ggagatttgg gcgtgcccc    240
gcaagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtgttact gcctgatagg    300
gtgcttgcca gtgccccggg aggtctcgta gaccgtgcac c                        341
```

<210> SEQ ID NO 50
 <211> LENGTH: 41
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 50

```
tgaacgggga gctaaccact ccaggccaat aggccttct g                        41
```

<210> SEQ ID NO 51
 <211> LENGTH: 42
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 51

```
tgaacgggga gctaaccact ccaggcctta agccatttcc tg                        42
```

-continued

<210> SEQ ID NO 52
 <211> LENGTH: 43
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 52

tgaaggttgg ggtaaacact ccggcctctt aagccatttc ctg 43

<210> SEQ ID NO 53
 <211> LENGTH: 16
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 53

ggtggctcca tcttag 16

<210> SEQ ID NO 54
 <211> LENGTH: 19
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 54

aatggtggct ccatcttag 19

<210> SEQ ID NO 55
 <211> LENGTH: 82
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 55

ccctagtcac ggctagctgt gaaaggtccg tgagccgcat gactgcagag agtgctgata 60

ctggcctctc tgcagatcat gt 82

<210> SEQ ID NO 56
 <211> LENGTH: 40
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 56

aggttggggg aaacactccg gcctcttaag ccatttcctg 40

<210> SEQ ID NO 57
 <211> LENGTH: 81
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 57

tttttttttt tttttttttt ttttttttct tttttttttt ctttcctttc cttctttttt 60

tcctttcttt ttcccttctt t 81

<210> SEQ ID NO 58
 <211> LENGTH: 101
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 58

aatggtggct ccatcttagc cctagtcacg gctagctgtg aaaggtccgt gagccgcatg 60

actgcagaga gtgctgatac tggcctctct gcagatcatg t 101

<210> SEQ ID NO 59

-continued

<211> LENGTH: 127
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 59

tgaaggttgg ggtaaacact ccggcctctt aagccatttc ctgttttttt tttttttttt	60
tttttttttt tttttttttt tttctttcct ttccttcttt ttttcctttc tttttccctt	120
ctttaat	127

<210> SEQ ID NO 60
 <211> LENGTH: 183
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 60

tgaaggttgg ggtaaacact ccggcctctt aagccatttc ctgttttttt tttttttttt	60
tttttttttt tttttttttt tttctttcct ttccttcttt ttttcctttc tttttccctt	120
ctttaatggt ggctccatct tagccctagt cacggctagc tgtgaaaggt ccgtgagccg	180
cat	183

<210> SEQ ID NO 61
 <211> LENGTH: 52
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 61

tgagccgc at gactgcagag agtgtctgata ctggcctctc tgcagatcat gt	52
---	----

<210> SEQ ID NO 62
 <211> LENGTH: 105
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 62

tgaattggt ggctccatct tagccctagt cacggctagc tgtgaaaggt ccgtgagccg	60
catgactgca gagagtgtctg atactggcct ctctgcagat catgt	105

<210> SEQ ID NO 63
 <211> LENGTH: 176
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 63

tgaaggttgg ggtaaacact ccggcctctt aagccatttc ctgttttttt tttttttttt	60
tttttttttt tttttttttt tttctttcct ttccttcttt ttttcctttc tttttccctt	120
ctttaatgcc gcatgactgc agagagtgtc gatactggcc tctctgcaga tcatgt	176

<210> SEQ ID NO 64
 <211> LENGTH: 200
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 64

tgacttaagc cttttcctgt tttttttttt tttttttttt tttttttcct tttttttttc	60
tttcctttcc ttcttttttt cctttctttt tcccttcttt aatgggtggc ccatcttagc	120
cctagtcaag gctagctgtg aaaggtccgt gagccgcatg actgcagaga gtgctgatac	180
tggcctctct gcagatcatg	200

-continued

<210> SEQ ID NO 65
 <211> LENGTH: 144
 <212> TYPE: DNA
 <213> ORGANISM: Hepatitis C virus

<400> SEQUENCE: 65

```
tgaaggttg ggtaaacact cgggcctctt aagccatttc ctgaatgggtg gctccatctt    60
agccctagtc acggctagct gtgaaaggtc cgtgagccgc atgactgcag agagtgctga    120
tactggcctc tctgcagatc atgt                                           144
```

What is claimed is:

1. A composition comprising a purified and isolated nucleic acid molecule, which encodes a chimeric human hepatitis C virus (HCV);

wherein expression of said nucleic acid molecule when operably linked to a promoter and transfected into cells results in production of the chimeric HCV; and wherein said nucleic acid molecule is producible by a process comprising the step of:

beginning with a HCV1a genomic nucleic acid encoding SEQ ID NO: 1, replacing the nucleic acid sequence that encodes at least one structural HCV1a protein with a the nucleic acid sequence that encodes the corresponding at least one structural HCV protein from the genome of an HCV from a genotype or a subtype different from HCV1a, to produce said nucleic acid molecule.

2. The composition according to claim 1, wherein said HCV1a genomic nucleic acid comprises SEQ ID NO:2.

3. The composition according to claim 1, wherein the process further comprises the step of inserting said nucleic acid molecule encoding the chimeric HCV into a cassette vector having the promoter upstream of the insertion site of said nucleic acid molecule.

4. The composition according to claim 3, wherein the cassette vector is pCV.

5. The composition according to claim 3, wherein the cassette vector contains one or more expressible marker genes.

6. The composition according to claim 3, wherein the promoter is a eukaryotic promoter.

7. The composition according to claim 3, wherein the promoter is a bacterial promoter.

8. The composition according to claim 3, wherein the promoter is a bacteriophage promoter.

9. The composition according to claim 8, wherein the bacteriophage promoter is a bacteriophage T7 promoter.

10. The composition according to claim 1, wherein the process further comprises the step of inserting said nucleic acid molecule encoding the chimeric HCV into a DNA construct capable of producing an RNA transcript of said nucleic acid molecule.

11. The composition according to claim 1, further comprising the chimeric HCV encoded by said nucleic acid molecule.

12. The composition according to claim 1, wherein the different genotype or the different serotype is HCV1b.

13. The composition according to claim 12, wherein said nucleic acid molecule comprises SEQ ID NO:6.

14. The composition according to claim 12, wherein said nucleic acid molecule encodes the amino acid sequence of SEQ ID NO:5.

15. A composition comprising a purified and isolated nucleic acid molecule, which encodes a chimeric human HCV;

wherein expression of said nucleic acid molecule when operably linked to a promoter and transfected into cells results in production of the chimeric HCV; and wherein said nucleic acid molecule is producible by a process comprising the step of:

beginning with a HCV1a genomic nucleic acid encoding SEQ ID NO: 1, replacing the nucleic acid sequence that encodes at least one nonstructural HCV1a protein with the nucleic acid sequence that encodes the corresponding at least one nonstructural HCV protein from the genome of an HCV from a genotype or a subtype different from HCV1a, to produce said nucleic acid molecule.

16. The composition of claim 15, wherein the nonstructural HCV protein is selected from the group consisting of NS3 protease, E1, E2 and NS4 proteins.

17. The composition according to claim 15, wherein said HCV1a genomic nucleic acid comprises SEQ ID NO:2.

18. The composition according to claim 15, wherein the process further comprises the step of inserting said nucleic acid molecule encoding the chimeric HCV into a cassette vector having the promoter upstream of the insertion site of said nucleic acid molecule.

19. The composition according to claim 18, wherein the cassette vector is pCV.

20. The composition according to claim 18, wherein the cassette vector contains one or more expressible marker genes.

21. The composition according to claim 18, wherein the promoter is a eukaryotic promoter.

22. The composition according to claim 18, wherein the promoter is a bacterial promoter.

23. The composition according to claim 18, wherein the promoter is a bacteriophage promoter.

24. The composition according to claim 23, wherein the bacteriophage promoter is a bacteriophage T7 promoter.

25. The composition according to claim 15, wherein the process further comprises the step of inserting said nucleic acid molecule encoding the chimeric HCV into a DNA construct capable of producing an RNA transcript of said nucleic acid molecule.

26. The composition according to claim 15, further comprising the chimeric HCV encoded by said nucleic acid molecule.

27. The composition according to claim 15, wherein said different genotype or serotype is HCV1b.

121

28. A method for inducing an immune response comprising administering to a mammal an effective amount of the composition according to claim **1**, **15**, **2-10**, **11**, **12-25**, or **26** to induce the immune response.

29. The method according to claim **28**, wherein the mammal is not infected with a hepatitis C virus. 5

30. The method according to claim **28**, wherein the mammal is infected with a hepatitis C virus.

31. A method for producing a chimeric human HCV molecule comprising the step of: 10

beginning with a HCV1a genomic nucleic acid encoding SEQ ID NO: 1, replacing the nucleic acid sequence that encodes at least one structural HCV1a protein with the nucleic acid sequence that encodes the corresponding at least one structural HCV protein from the genome of an HCV from a genotype or a subtype different from HCV1a, to produce said chimeric HCV molecule; 15

wherein expression of said chimeric HCV molecule when operably linked to a promoter and transfected into cells results in production of the chimeric HCV. 20

32. A method for producing a chimeric human HCV molecule comprising the step of:

beginning with a HCV1a genomic nucleic acid encoding SEQ ID NO: 1, replacing the nucleic acid sequence that encodes at least one nonstructural HCV protein with the nucleic acid sequence that encodes the corresponding at least one nonstructural HCV protein from the genome of an HCV from a genotype or a subtype different from HCV1a, to produce said chimeric HCV molecule; 25

wherein expression of said chimeric HCV molecule when operably linked to a promoter and transfected into cells results in production of the chimeric HCV. 30

33. The method according to claim **32**, wherein the non-structural HCV protein is selected from the group consisting of NS3 protease, E1, E2 and NS4 proteins.

122

34. A composition comprising a purified and isolated nucleic acid molecule, which encodes a chimeric human HCV;

wherein expression of said nucleic acid molecule when operably linked to a promoter and transfected into cells results in production of the chimeric HCV;

wherein said nucleic acid molecule is producible by a process comprising the step of:

beginning with a HCV1b genomic nucleic acid encoding SEQ ID NO:3, replacing the nucleic acid sequence that encodes at least one HCV1b structural protein with the nucleic acid sequence that encodes the corresponding at least one structural protein from the genome of an HCV from a genotype or a subtype different from HCV1b, to produce said nucleic acid molecule.

35. The composition according to claim **34**, wherein said HCV1b genomic nucleic acid comprises SEQ ID NO:4.

36. A composition comprising a purified and isolated nucleic acid molecule, which encodes a chimeric human HCV, 20

wherein expression of said nucleic acid molecule when operably linked to a promoter and transfected into cells results in production of the chimeric HCV,

wherein said nucleic acid molecule is producible by a process comprising the step of:

beginning with a HCV1b genomic nucleic acid encoding SEQ ID NO:3, replacing the nucleic acid sequence that encodes at least one nonstructural HCV1b protein with the nucleic acid sequence that encodes the corresponding at least one nonstructural protein from the genome of an HCV from a genotype or a subtype different from HCV1b, to produce said nucleic acid molecule.

37. The composition according to claim **36**, wherein said HCV1b genomic nucleic acid comprises SEQ ID NO:4.

* * * * *